UNIVERSITY OF CALIFORNIA

Los Angeles

Diffractive Optical Networks

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Electrical and Computer Engineering

by

Deniz Mengu

2022

ABSTRACT OF THE DISSERTATION

Diffractive Optical Neural Networks

by

Deniz Mengu

Doctor of Philosophy in Electrical and Computer Engineering

University of California, Los Angeles, 2022

Professor Aydogan Ozcan, Chair

Deep learning has been revolutionizing information processing in many fields of science and engineering owing to the massively growing amounts of data and the advances in deep neural network architectures. As these neural networks are expanding their capabilities towards achieving state-of-the-art solutions for demanding statistical inference tasks in various applications, there appears to be a global need for low-power, scalable and fast computing hardware beyond what existing electronic systems can offer. Optical computing might potentially address some of these demands with its inherent parallelism, power efficiency, and high speed. Recent advances in optical materials, fabrication, and optimization techniques have significantly enriched the design capabilities in optics and photonics, leading to various successful demonstrations of guided-wave and free-space computing hardware for accelerating machine learning tasks using light. While integrated waveguide-based photonic approaches

mainly aims to replace the current electronic computing hardware with better alternatives, free-space optical neural network architectures and related computing techniques offer unique advantages particularly for inference tasks in visual computing applications, where the information is already in the optical domain.

This dissertation introduces diffractive optical networks that are designed based on Diffractive Deep Neural Networks (D^2NN) framework using deep learning to tackle various challenges in computational machine vision by providing power-efficient, fast, scalable and massively parallel all-optical solutions. First, a series of design advances were devised to improve the statistical inference accuracy of diffractive object classifiers. Second, hybrid (optical-electronic) neural network systems, which uses diffractive optical networks as front-end optical processors preceding back-end electronic neural networks, were investigated to enable task-specific camera systems that can perform object classification with fewer pixels, thus with less memory and power consumption. In addition, D^2NN framework was extended to mitigate the adverse impact of possible physical error sources, termed as vaccinated-D^2NN (v-D^2NN). The success of v-D^2NN was experimentally demonstrated at THz wavelengths by comparing the classification accuracies of 3D-printed nonvaccinated and vaccinated diffractive handwritten digit classifiers under the presence of layer-to-layer misalignments. Next, a diffractive all-optical object classifier was designed to provide inference accuracy that is invariant under random changes on the scale, position and orientation of the input objects with respect to the diffractive surfaces. Furthermore, the all-optical information processing capacity of diffractive optical networks was studied to prove that the dimensionality of the solution space representing the set of all-optical transformations established by a diffractive network increases linearly with the number of diffractive surfaces, up to a limit determined by the size of the input/output fields-of-

view. In parallel, the diffractive optical networks were shown to all-optically perform arbitrary complex-valued linear transformations, including space-variant operations, noninvertible and nonunitary matrices, with negligibly small errors provided that the total number of diffractive neurons is sufficiently large to satisfy space-bandwidth product demands on input and output fields-of-view. A diffractive permutation network that can all-optically implement 625 interconnects between its input and output was fabricated using 3D printing and its performance was demonstrated at THz wavelengths.

Beyond the outlined optical computing and machine learning applications, diffractive optical networks can also be utilized to all-optically solve challenging inverse problems in computational imaging. Highlighting this aspect, diffractive optical networks that can all-optically perform phase retrieval to reveal the quantitative phase image (QPI) of weakly scattering objects were devised. Based on the conducted analysis, these diffractive QPI networks can resolve subwavelength features, ~0.67λ, of an input phase object, with λ denoting the wavelength of illumination. Finally, in certain application scenarios, spatial overlap between phase objects poses an irreversible information loss due to the superposition of individual phase delays. It was demonstrated that diffractive optical networks can be trained to solve this challenging problem to infer the classes of spatially overlapping phase objects. Moreover, when these diffractive phase object classifiers are combined with electronic deep neural networks, the individual phase images of the objects spatially overlapping within the input field-of-view can be recovered based on the all-optically synthesized class scores, despite the phase ambiguity.

All the studies presented in this dissertation demonstrating the success of diffractive optical networks in various general-purpose computing, statistical inference and inverse computational

imaging tasks can potentially lead them to largely replace conventional optical components in the next-generation, task-specific machine vision designs that can achieve a given task with fewer pixels, leading to faster, more memory- and power-efficient systems.

The dissertation of Deniz Mengu is approved.

Mona Jarrahi

Dino Di Carlo

Chee Wei Wong

Yair Rivenson

Aydogan Ozcan, Committee Chair

University of California, Los Angeles

2022

To my mom, *Ozden*, and dad, *Derya*

Table of Contents

List of Figures

List of Tables

Acknowledgements

First, I would like express to my sincere gratitude to Prof. Aydogan Ozcan. His enthusiasm, always measured and calculated mentorship along with his support were invaluable. He certainly expanded my horizons as a scientist and an engineer beyond what I had ever imagined. His unique vision and unprecedented communication skills has been extremely inspiring for me. Beyond science and engineering, he also challenged my professionalism, work ethic and durability in a way that had never been challenged before. I had to make huge compromises on the other aspects of my life to keep up with the pace of the progress and research Prof. Ozcan demands, which, in hindsight, were worth to maximize the output and productivity during my time under his guidance and supervision, because the opportunity of working with him is a lifechanging experience and an incredible privilege only a handful of very fortunate people can get.

I would like to thank to Prof. Mona Jarrahi. Having access to her expertise, vision and leadership through our collaboration has been crucial in the progress of my PhD studies. I would like to thank to Dr. Yair Rivenson for his support and guidance. I feel fortunate to be able to have fruitful scientific discussions as well as enlightening non-scientific conversations with him, and to some extent those discussions provided the mental stimulation that I desperately needed beyond the never-ending, intellectually restricted/confined progress-presentation cycles. I also would like to thank to my doctoral committee members, Prof. Dino Di Carlo and Prof. Chee Wei Wong for their kind support during my PhD.

I would like to also thank Prof. Hakan Urey and Dr. Erdem Ulusoy, who I had the honor to work with during my MSc. studies as a member of Optical Microsystems Laboratory in Koc

University, Istanbul, Turkey. This PhD thesis is not only a result of my hardwork, Prof. Ozcan's leadership, creativity and academic excellence, but it is also the outcome of the solid scientific foundation that I managed to establish during my masters under their impeccable guidance.

I would like to thank to my colleagues, friends and co-authors who tremendously helped me during this journey: Muhammed Veli, Nezih Tolga Yardimci, Hatice Ceylan Koydemir, Onur Kulce, Yi Luo, Cagatay Isil, Xurong Li, Hyou-arm Joung, Zoltan Gorocs, Zach Ballard, Calvin Brown, Mustafa Ugur Daloglu, Derek Tseng, Yibo Zhang, Hongda Wang, Artem Goncharov, Jingxi Li, Sadman Rahman, Yifan Zhao, Ani Ray, Yichen Wu, and my favorite married couple Deniz Turan and Eylul Simsek, I have immensely enjoyed working, in the case of Deniz-Eylul living, in such a collaborative and rich community of engineers, scientists and friends.

I would like to thank to my family for their great and unconditional support in my journey. Last but not least, I would like to also thank Gizem Guzelsoy, soon to be a PhD, for her huge support and patience while I have been living like a lab mouse running through the repeating and endless cycles of progress meetings, presentations and publications. She has been the greatest source of joy in my life throughout my PhD studies. Without her undying companionship and camaraderie, I am not sure if I could pass the finish line before having a mental breakdown.

Vita

Deniz Mengu received his B.Sc. degree in electrical and electronics engineering from Middle East Technical University, Ankara, Turkey. He got his M.Sc. degree in electrical and electronics engineering from Koç University, Istanbul, Turkey, working under the supervision of Prof. Hakan Urey and Dr. Erdem Ulusoy. He, then, joined the Bio- and Nano-photonics Lab led by Prof. Aydogan Ozcan at UCLA in Fall 2016. He has co-authored 18 journal articles, various conference proceedings and several US patents on computational imaging/sensing, holographic near-eye displays, optical machine learning and computing.

Selected Publications

1. **Mengu D**, Luo Y, Rivenson Y, Ozcan A. Analysis of diffractive optical neural networks and their integration with electronic neural networks. IEEE Journal of Selected Topics in Quantum Electronics. 2019 Jun 6;26(1):1-4. DOI: 10.1109/JSTQE.2019.2921376

2. **Mengu D**, Zhao Y, Yardimci NT, Rivenson Y, Jarrahi M, Ozcan A. Misalignment resilient diffractive optical networks. Nanophotonics. 2020 Oct 1;9(13):4207-19. DOI: 10.1515/nanoph-2020-0291

3. **Mengu D**, Rivenson Y, Ozcan A. Scale-, shift-, and rotation-invariant diffractive optical networks. ACS photonics. 2020 Dec 23;8(1):324-34. DOI: 10.1021/acsphotonics.0c01583

4. Kulce O, **Mengu D**, Rivenson Y, Ozcan A. All-optical information-processing capacity of diffractive surfaces. Light: Science & Applications. 2021 Jan 28;10(1):1-7. DOI: 10.1038/s41377-020-00439-9

5. Kulce O, **Mengu D**, Rivenson Y, Ozcan A. All-optical synthesis of an arbitrary linear transformation using diffractive surfaces. Light: Science & Applications. 2021 Sep 24;10(1):1-21. DOI: 10.1038/s41377-021-00623-5

6. **Mengu D**, Rahman MS, Luo Y, Li J, Kulce O, Ozcan A. At the intersection of optics and deep learning: statistical inference, computing, and inverse design. Advances in Optics and Photonics. 2022 Jun 30;14(2):209-90. DOI: 10.1364/AOP.450345

7. **Mengu D**, Ozcan A. All□Optical Phase Recovery: Diffractive Computing for Quantitative Phase Imaging. Advanced Optical Materials. 2022:2200281. DOI: 10.1002/adom.202200281

8. **Mengu D**, Veli M, Rivenson Y, Ozcan A. Classification and reconstruction of spatially overlapping phase images using diffractive optical networks. Scientific reports. 2022 May 19;12(1):1-8. DOI: 10.1038/s41598-022-12020-y

9. **Mengu D**, Zhao Y, Tabassum A, Jarrahi M, Ozcan A. Diffractive interconnects: all-optical permutation operation using diffractive networks. Nanophotonics. 2022 Sep 5. DOI: 10.1515/nanoph-2022-0358

10. Li J, **Mengu D**, Yardimci NT, Luo Y, Li X, Veli M, Rivenson Y, Jarrahi M, Ozcan A. Spectrally encoded single-pixel machine vision using diffractive networks. Science Advances. 2021 Mar 26;7(13). DOI: 10.1126/sciadv.abd769012.

11. Veli M, **Mengu D**, Yardimci NT, Luo Y, Li J, Rivenson Y, Jarrahi M, Ozcan A. Terahertz pulse shaping using diffractive surfaces. Nature Communications. 2021 Jan 4;12(1):1-3. DOI: 10.1038/s41467-020-20268-z

Chapter 1 Analysis of Diffractive Optical Neural Networks and Their Integration With Electronic Neural Networks

Parts of this chapter have previously been published in D. Mengu et al. “Analysis of Diffractive Optical Neural Networks and Their Integration With Electronic Neural Networks”. IEEE JSTQE, 2020, 26(1), 1–14, 3700114, DOI: 10.1109/JSTQE.2019.2921376.

Optical machine learning offers advantages in terms of power efficiency, scalability, and computation speed. Recently, an optical machine learning method based on diffractive deep neural networks (D^2NNs) has been introduced to execute a function as the input light diffracts through passive layers, designed by deep learning using a computer. In this chapter, I introduce improvements to D^2NNs by changing the training loss function and reducing the impact of vanishing gradients in the error back-propagation step. As a result of these design advances, based on five phase-only diffractive layers, the reported diffractive optical networks can numerically achieve a classification accuracy of 97.18% and 89.13% for optical recognition of handwritten digits and fashion products, respectively; using both phase and amplitude modulation (complex-valued) at each layer, the inference performance improves to 97.81% and 89.32%, respectively. Furthermore, this chapter reports the integration of D^2NNs with electronic neural networks to create hybrid classifiers that significantly reduce the number of input pixels into an electronic network using an ultra-compact front-end D^2NN with a layer-to-layer distance of a few wavelengths, also reducing the complexity of the successive electronic network. Using a five-layer phase-only D^2NN jointly optimized with a single fully connected electronic layer, the hybrid neural network system achieves a classification accuracy of 98.71% and 90.04% for the recognition of handwritten digits and fashion products, respectively, despite the signal is

compressed by >7.8 times down to 10 × 10 pixels due to limited space-bandwidth product of the focal plane-array. Beyond creating low-power and high-frame rate machine learning platforms, D^2NN-based hybrid neural networks will find applications in smart optical imager and sensor design.

1.1 Introduction

Optics in machine learning has been widely explored due to its unique advantages, encompassing power efficiency, speed and scalability[1–3]. Some of the earlier work include optical implementations of various neural network architectures[4–10], with a recent resurgence[11–22], following the availability of powerful new tools for applying deep neural networks[23,24], which have redefined the state-of-the-art for a variety of machine learning tasks. In this line of work, an optical machine learning framework has recently been developed and introduced, termed as Diffractive Deep Neural Network (D^2NN)[15], where deep learning and error back-propagation methods are used to design, inside a computer, diffractive layers that collectively perform a desired task that the network is trained for. In this training phase of a D^2NN, the transmission and/or reflection coefficients of the individual pixels (i.e., neurons) of each layer are optimized such that as the light diffracts from the input plane toward the output plane, it computes the task at hand. Once this training phase in a computer is complete, these passive layers can be physically fabricated and stacked together to form an all-optical network that executes the trained function without the use of any power, except for the illumination light and the output detectors.

In our previous work, we experimentally demonstrated the success of D^2NN framework at THz part of the electromagnetic spectrum and used a standard 3D-printer to fabricate and

assemble together the designed D^2NN layers[15]. In addition to demonstrating optical classifiers, we also demonstrated that the same D^2NN framework can be used to design an imaging system by 3D-engineering of optical components using deep learning[15]. In these earlier results, we used coherent illumination and encoded the input information in phase or amplitude channels of different D^2NN systems. Another important feature of D^2NNs is that the axial spacing between the diffractive layers is very small, e.g., less than 50 wavelengths (λ)[15], which makes the entire design highly compact and flat.

Our experimental demonstration of D^2NNs was based on linear materials, without including the equivalent of a nonlinear activation function within the optical network; however, as detailed in [15], optical nonlinearities can also be incorporated into a D^2NN using non-linear materials including e.g., crystals, polymers or semiconductors, to potentially improve its inference performance using nonlinear optical effects within diffractive layers. For such a nonlinear D^2NN design, resonant nonlinear structures (based on e.g., plasmonics or metamaterials) tuned to the illumination wavelength could be important to lower the required intensity levels. Even using linear optical materials to create a D^2NN, the optical network designed by deep learning shows "*depth*" advantage, i.e., a single diffractive layer does not possess the same degrees-of-freedom to achieve the same level of classification accuracy, power efficiency and signal contrast at the output plane that multiple diffractive layers can collectively achieve for a given task. It is true that, for a linear diffractive optical network, the entire wave propagation and diffraction phenomena that happen between the input and output planes can be squeezed into a *single* matrix operation; *however*, this arbitrary mathematical operation defined by multiple learnable diffractive layers cannot be performed in general by a single diffractive layer placed between the same input and output planes. That is why, multiple diffractive layers forming a D^2NN show the

depth advantage, and statistically perform better compared to a single diffractive layer trained for the same classification task, and achieve improved accuracy as also discussed in the supplementary materials of [15].

Here, we present a detailed analysis of D^2NN framework, covering different parameters of its design space, also investigating the advantages of using multiple diffractive layers, and provide significant improvements to its inference performance by changing the loss function involved in the training phase, and reducing the effect of vanishing gradients in the error back-propagation step through its layers. To provide examples of its improved inference performance, using a 5-layer D^2NN design (Fig. 1.1), we optimized two different classifiers to recognize (1) hand-written digits, 0 through 9, using the MNIST (Mixed National Institute of Standards and Technology) image dataset[25], and (2) various fashion products, including t-shirts, trousers, pullovers, dresses, coats, sandals, shirts, sneakers, bags, and ankle boots (using the Fashion MNIST image dataset[26]). These 5-layer phase-only all-optical diffractive networks achieved a numerical blind testing accuracy of 97.18% and 89.13% for hand-written digit classification and fashion product classification, respectively. Using the same D^2NN design, this time with both the phase and the amplitude of each neuron's transmission as learnable parameters (which we refer to as *complex-valued* D^2NN design), we improved the inference performance to 97.81% and 89.32% for hand-written digit classification and fashion product classification, respectively. We also provide comparative analysis of D^2NN performance as a function of our design parameters, covering the impact of the number of layers, layer-to-layer connectivity and loss function used in the training phase on the overall classification accuracy, output signal contrast and power efficiency of D^2NN framework.

Furthermore, we report the integration of D^2NNs with electronic neural networks to create hybrid machine learning and computer vision systems. Such a hybrid system utilizes a D^2NN at its front-end, before the electronic neural network, and if it is jointly optimized (i.e., optical and electronic as a monolithic system design), it presents several important advantages. This D^2NN-based hybrid approach can all-optically *compress* the needed information by the electronic network using a D^2NN at its front-end, which can then significantly reduce the number of pixels (detectors) that needs to be digitized for an electronic neural network to act on. This would further improve the frame-rate of the entire system, also reducing the complexity of the electronic network and its power consumption. This D^2NN-based hybrid design concept can potentially create ubiquitous and low-power machine learning systems that can be realized using relatively simple and compact imagers, with e.g., a few tens to hundreds of pixels at the opto-electronic sensor plane, preceded by an ultra-compact all-optical diffractive network with a layer-to-layer distance of a few wavelengths, which presents important advantages compared to some other hybrid network configurations involving e.g., a 4-f configuration[16] to perform a convolution operation before an electronic neural network.

To better highlight these unique opportunities enabled by D^2NN-based hybrid network design, we conducted an analysis to reveal that a 5-layer phase-only (or *complex-valued*) D^2NN that is jointly-optimized with a single fully-connected layer, following the optical diffractive layers, achieves a blind classification accuracy of 98.71% (or *98.29%*) and 90.04% (or *89.96%*) for the recognition of hand-written digits and fashion products, respectively. In these results, the input image to the electronic network (created by diffraction through the jointly-optimized front-end D^2NN) was also compressed by more than 7.8 times, down to 10×10 pixels, which confirms that a D^2NN-based hybrid system can perform competitive classification performance even using

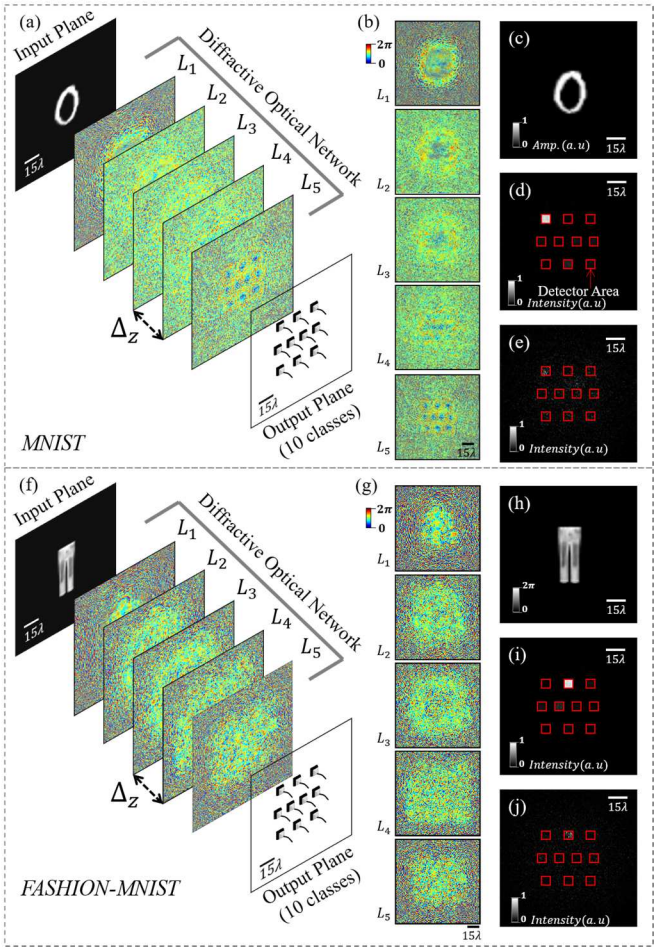

Fig. 1.1 All-optical diffractive classifier networks. These D^2NN designs were based on spatially and temporally coherent illumination and linear optical materials/layers. (a) D^2NN setup for the task of classification of handwritten digits (MNIST), where the input information is encoded in the *amplitude* channel of the input plane. (b) Final design of a 5-layer, phase-only classifier for handwritten digits. (c) Amplitude distribution at the input plane for a test sample (digit '0'). (d-e) Intensity patterns at the output plane for the input in (c); (d) is for MSE-based, and (e) is softmax-cross-entropy (SCE)-based designs. (f) D^2NN setup for the task of classification of fashion products (Fashion-MNIST), where the input information is encoded in the *phase* channel of the input plane. (g) Same as (b), except for fashion product dataset. (h) Phase distribution at the input plane for a test sample. (i-j) Same as (d) and (e) for the input in (h). λ refers to the illumination source wavelength. Input plane represents the plane of the input object or its data, which can also be generated by another optical imaging system or a lens, projecting an image of the object data onto this plane.

a

relatively simple and one-layer electronic network that uses significantly reduced number of input pixels.

In addition to potentially enabling ubiquitous, low-power and high-frame rate machine learning and computer vision platforms, these hybrid neural networks which utilize D^{2}NN-based all-optical processing at its front-end will find other applications in the design of compact and ultra-thin optical imaging and sensing systems by merging fabricated D^{2}NNs with opto-electronic sensor arrays. This will create intelligent systems benefiting from various CMOS/CCD imager chips and focal plane arrays at different parts of the electromagnetic spectrum, merging the benefits of all-optical computation with simple and low-power electronic neural networks that can work with lower dimensional data, all-optically generated at the output of a jointly-optimized D^{2}NN design.

1.2 Results and Discussion

Mitigating vanishing gradients in optical neural network training

In D^{2}NN framework, each neuron has a complex transmission coefficient, i.e., $t_i^l(x_i, y_i, z_i) = a_i^l(x_i, y_i, z_i)exp(j\phi_i^l(x_i, y_i, z_i))$, where i and l denote the neuron and diffractive layer number, respectively. In [15], a_i^l and ϕ_i^l are represented during the network training as functions of two latent variables, α and β, defined in the following form:

$$a_i^l = sigmoid(\alpha_i^l),$$
$$\phi_i^l = 2\pi \times sigmoid(\beta_i^l), \quad (1.1)$$

where, $sigmoid(x) = \frac{e^x}{e^x+1}$, is a non-linear, differentiable function. In fact, the trainable

parameters of a D^{2}NN are these latent variables, α_i^l and β_i^l, and Eq. (1.1) defines how they are related to the physical parameters (a_i^l and ϕ_i^l) of a diffractive optical network. Note that in Eq. (1.1), the sigmoid acts on an auxiliary variable rather than the information flowing through the network. Being a bounded analytical function, sigmoid confines the values of a_i^l and ϕ_i^l inside the intervals $(0,1)$ and $(0,2\pi)$, respectively. On the other hand, it is known that sigmoid function has vanishing gradient problem[27] due to its relatively flat tails, and when it is used in the context depicted in Eq. (1.1), it can prevent the network to utilize the available dynamic range considering both the amplitude and phase terms of each neuron. To mitigate these issues, we replaced Eq. (1.1) as follows:

$$a_i^l = \frac{ReLU(\alpha_i^l)}{max_{0<i\leq M}\{ReLU(\alpha_i^l)\}},$$

$$\phi_i^l = 2\pi \times \beta_i^l, \tag{1.2}$$

where ReLU refers to Rectified Linear Unit, and M is the number of neurons per layer. Based on Eq. (1.2), the phase term of each neuron, ϕ_i^l, becomes unbounded, but since the $exp(j\phi_i^l(x_i, y_i, z_i))$ term is periodic (and bounded) with respect to ϕ_i^l, the error back-propagation algorithm is able to find a solution for the task in hand. The amplitude term, a_i^l, on the other hand, is kept within the interval $(0,1)$ by using an explicit normalization step shown in Eq. (1.2).

To exemplify the impact of this change *alone* in the training of an all-optical D^{2}NN design, for a 5-layer, phase-only (*complex-valued*) diffractive optical network with an axial distance of $40\times\lambda$ between its layers, the classification accuracy for Fashion-MNIST dataset increased from reported 81.13% (*86.33%*) to 85.40% (*86.68%*) following the above discussed changes in the

parameterized formulation of the neuron transmission values compared to earlier results in [15].We will report further improvements in the inference performance of an all-optical D^2NN after the introduction of the loss function related changes into the training phase, which is discussed next.

We should note that although the results of this paper follow the formulation in Eq. (1.2), it is also possible to parameterize complex modulation terms over the real and imaginary parts as in [28] and a formulation based on the Wirtinger derivatives can be used for error backpropagation.

Effect of the learning loss function on the performance of all-optical diffractive neural networks

Earlier work on D^2NNs[15] reports the use of mean squared error (MSE) loss. An alternative loss function that can be used for the design of a D^2NN is the cross-entropy loss[29,30] (see the Methods section). Since minimizing the cross-entropy loss is equivalent to minimizing the negative log-likelihood (or maximizing the likelihood) of an underlying probability distribution, it is in general more suitable for classification tasks. Note that, cross-entropy acts on probability measures, which take values in the interval $(0,1)$ and the signals coming from the detectors (one for each class) at the output layer of a D^2NN are not necessarily in this range; therefore, in the training phase, a *softmax* layer is introduced to be able to use the cross-entropy loss. It is important to note that although *softmax* is used during the *training* process of a D^2NN, once the diffractive design converges and is fixed, the class assignment at the output plane of a D^2NN is still based *solely on the maximum optical signal detected at the output plane*, where there is one detector assigned for each class of the input data (see Figs. 1.1(a), 1.1(f)).

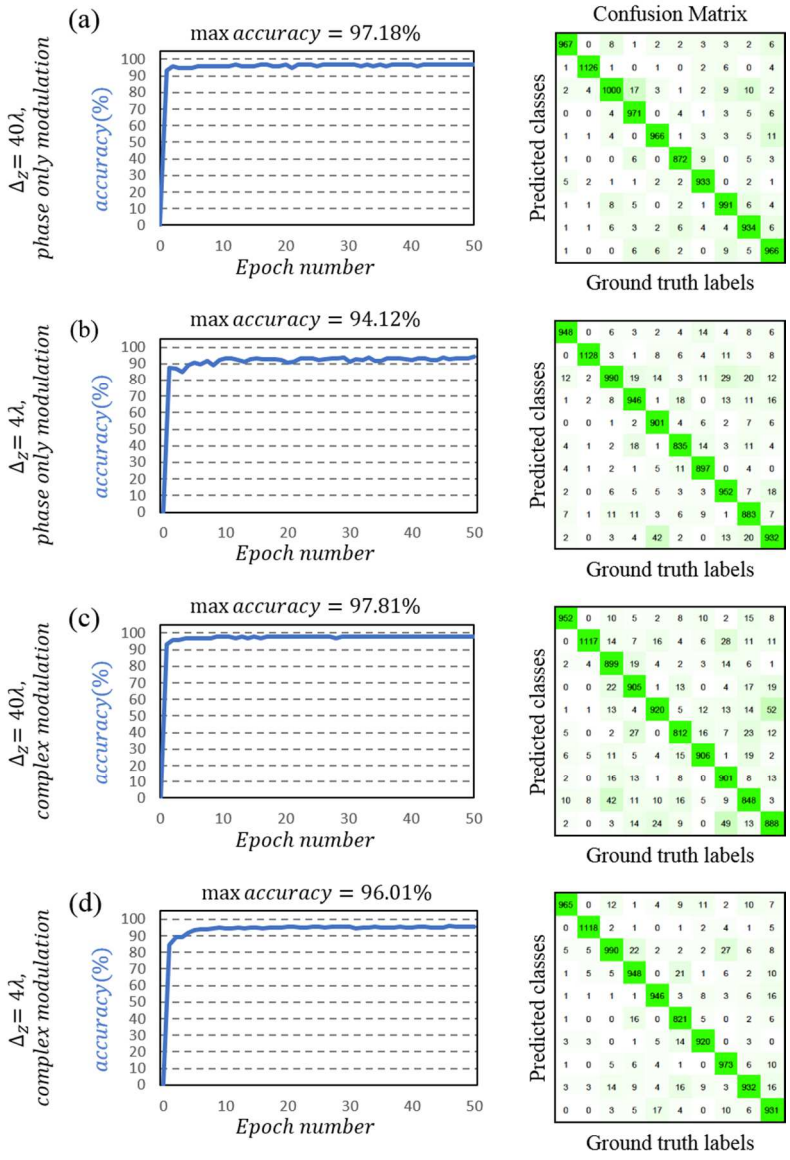

Fig. 1.2 Convergence plots and confusion matrices for all-optical D2NN-based classification of handwritten digits (MNIST dataset). (a) Convergence curve and confusion matrix for a phase-only, fully-connected D2NN ($\Delta_z = 40\lambda$) design. (b) Convergence curve and confusion matrix for a phase-only, partially-connected D2NN ($\Delta_z = 4\lambda$) design. (c) and (d) are

When we combine D^{2}NN training related changes reported in the earlier sub-section on the parametrization of neuron modulation (Eq. (1.2)), with the cross-entropy loss outlined above, a

significant improvement in the classification performance of an all-optical diffractive neural network is achieved. For example, for the case of a 5-layer, phase-only D^2NN with $40\times\lambda$ axial distance between the layers, the classification accuracy for MNIST dataset increased from 91.75% to 97.18%, which further increased to 97.81% using complex-valued modulation, treating the phase and amplitude coefficients of each neuron as learnable parameters. The training convergence plots and the confusion matrices corresponding to these results are also reported in Figs. 1.2(a) and 1.2(c), for phase-only and complex-valued modulation cases, respectively. Similarly, for Fashion-MNIST dataset, we improved the blind testing classification accuracy of a 5-layer phase-only (*complex-valued*) D^2NN from 81.13% (*86.33%*) to 89.13% (*89.32%*), showing a similar level of advancement as in the MNIST results. Figs. 1.3(a) and 1.3(c) also report the training convergence plots and the confusion matrices for these improved Fashion-MNIST inference results, for phase-only and complex-valued modulation cases, respectively. As a comparison point, a fully-electronic deep neural network such as ResNet-50[31] (with >25 Million learnable parameters) achieves 99.51% and 93.23% for MNIST and Fashion-MNIST datasets, respectively, which are superior to our 5-layer all-optical D^2NN inference results (i.e., 97.81% and 89.32% for MNIST and Fashion-MNIST datasets, respectively), which in total used 0.8 million learnable parameters, covering the phase and amplitude values of the neurons at 5 successive diffractive layers.

All these results demonstrate that the D^2NN framework using linear optical materials can already achieve a decent classification performance, also highlighting the importance of future research on the integration of optical nonlinearities into the layers of a D^2NN, using e.g., plasmonics, metamaterials or other nonlinear optical materials (see the supplementary information of [15]), in order to come closer to the performance of state-of-the-art digital deep

neural networks.

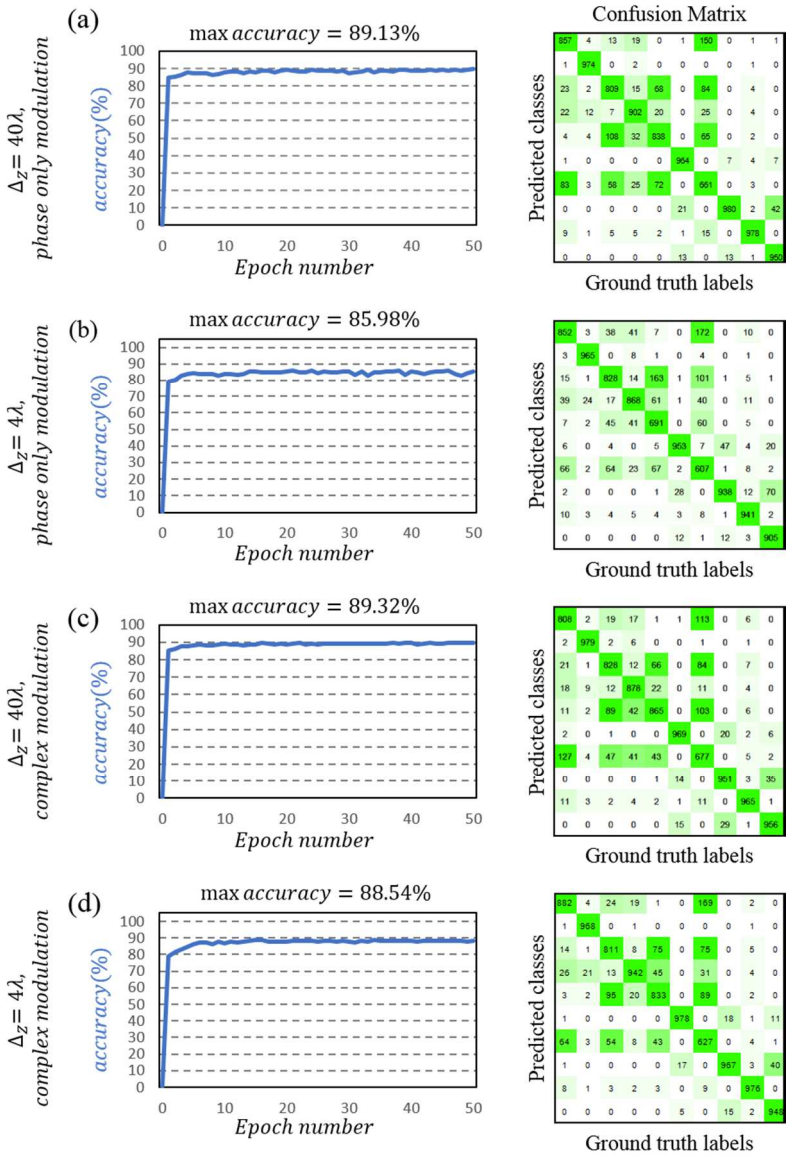

Fig. 1.3 Same as Fig. 1.2, except the results are for all-optical D2NN-based classification of fashion products (Fashion-MNIST dataset) encoded in the phase channel of the input plane.

Performance trade-offs in D^{2}NN design

Despite the significant increase observed in the blind testing accuracy of D^{2}NNs, the use of softmax-cross-entropy (SCE) loss function in the context of all-optical networks also presents some trade-offs in terms of practical system parameters. MSE loss function operates based on pixel-by-pixel comparison of a user-designed output distribution with the output optical intensity pattern, after the input light interacts with the diffractive layers (see e.g., Figs. 1.1(d) and 1.1(i)). On the other hand, SCE loss function is much less restrictive for the spatial distribution or the uniformity of the output intensity at a given detector behind the diffractive layers (see e.g., Figs. 1.1(e) and 1.1(j)); therefore, it presents additional degrees-of-freedom and redundancy for the diffractive network to improve its inference accuracy for a given machine learning task, as reported in the earlier sub-section.

This performance improvement with the use of SCE loss function in a diffractive neural network design comes at the expense of some compromises in terms of the expected diffracted power efficiency and signal contrast at the network output. To shed more light on this trade-off, we define the power efficiency of a D^{2}NN as the percentage of the optical signal detected at the target label detector (I_L) corresponding to the correct data class with respect to the *total* optical signal at the output plane of the optical network (E). Fig. 1.4(b) and Fig. 1.4(e) show the power efficiency comparison as a function of the number of diffractive layers (corresponding to 1, 3 and 5-layer phase-only D^{2}NN designs) for MNIST and Fashion-MNIST datasets, respectively. The power efficiency values in these graphs were computed as the ratio of the mean values of I_L and E for the test samples that were correctly classified by the corresponding D^{2}NN designs (refer to Figs. 1.4(a) and 1.4(d) for the classification accuracy of each design). These results

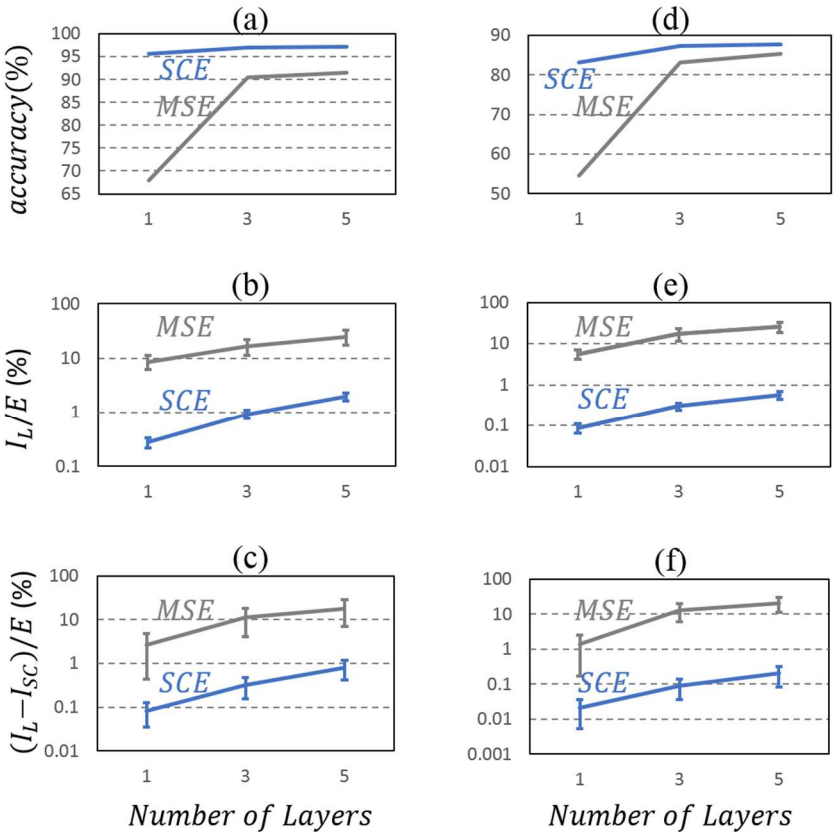

Fig. 1.4 Classification accuracy, power efficiency and signal contrast comparison of MSE and SCE loss function based all-optical phase-only D2NN classifier designs with 1, 3 and 5-layers. (a) Blind testing accuracy, (b) power efficiency and (c) signal contrast analysis of the final design of fully-connected, phase-only all-optical classifiers trained for handwritten digits (MNIST). (d-f) are the same as (a-c), only the classified dataset is Fashion-MNIST instead.

clearly indicate that increasing the number of diffractive layers has significant positive impact on the optical efficiency of a D^2NN, regardless of the loss function choice. The maximum efficiency that a 5-layer phase-only D^2NN design based on the SCE loss function can achieve is 1.98% for MNIST and 0.56% for Fashion-MNIST datasets, which are significantly lower compared to the efficiency values that diffractive networks designed with MSE loss function can achieve, i.e., 25.07% for MNIST and 26.00% for Fashion-MNIST datasets (see Figs. 1.4(b) and 1.4(e)). Stated

differently, MSE loss function based D^2NNs are in general significantly more power efficient all-optical machine learning systems.

Next we analyzed the signal contrast of diffractive neural networks, which we defined as the difference between the optical signal captured by the target detector (I_L) corresponding to the correct data class and the maximum signal detected by the rest of the detectors (i.e., the strongest competitor (I_{SC}) detector for each test sample), normalized with respect to the total optical signal at the output plane (E). The results of our signal contrast analysis are reported in Fig. 1.4(c) and Fig. 1.4(f) for MNIST and Fashion-MNIST datasets, respectively, which reveal that D^2NNs designed with an MSE loss function keep a strong margin between the target detector (I_L) and the strongest competitor detector (among the rest of the detectors) at the output plane of the all-optical network. The minimum mean signal contrast value observed for an MSE-based D^2NN design was for a 1-Layer, phase-only diffractive design, showing a mean signal contrast of 2.58% and 1.37% for MNIST and Fashion-MNIST datasets, respectively. Changing the loss function to SCE lowers the overall signal contrast of diffractive neural networks as shown in Figs. 1.4(c) and 1.4(f).

Comparing the performances of MSE-based and SCE-based D^2NN designs in terms of classification accuracy, power efficiency and signal contrast, as depicted in Fig. 1.4, we identify two opposite design strategies in diffractive all-optical neural networks. MSE, being a strict loss function acting in the physical space (e.g., Figs. 1.1(d) and 1.1(i)), promotes high signal contrast and power efficiency of the diffractive system, while SCE, being much less restrictive in its output light distribution (e.g., Figs. 1.1(e) and 1.1(j)), enjoys more degrees-of-freedom to improve its inference performance for getting better classification accuracy, at the cost of a

reduced overall power efficiency and signal contrast at its output plane, which increases the systems' vulnerability for opto-electronic detection noise. In addition to the noise at the detectors, mechanical misalignment in both the axial and lateral directions might cause inference discrepancy between the final network model and its physical implementation. One way to mitigate this alignment issue is to follow the approach in Ref. [15] where the neuron size was chosen to be >3-4 times larger than the available fabrication resolution. Recently developed micro- and nano-fabrication techniques, such as laser lithography based on two-photon polymerization [32], emerge as promising candidates towards monolithic fabrication of complicated volumetric structures, which might help to minimize the alignment challenges in diffractive optical networks. Yet, another method of increasing the robustness against mechanical fabrication and related alignment errors is to model and include these error sources as part of the forward model during the numerical design phase, which might create diffractive models that are more tolerant of such errors.

Advantages of multiple diffractive layers in D^2NN framework

As demonstrated in Fig. 1.4, multiple diffractive layers that collectively operate within a D^2NN design present additional degrees-of-freedom compared to a single diffractive layer to achieve better classification accuracy, as well as improved diffraction efficiency and signal contrast at the output plane of the network; the latter two are especially important for experimental implementations of all-optical diffractive networks as they dictate the required illumination power levels as well as signal-to-noise ratio related error rates for all-optical classification tasks. Stated differently, D^2NN framework, even when it is composed of linear optical materials, shows depth advantage because an increase in the number of diffractive layers (1) improves its statistical inference accuracy (see Figs. 1.4(a) and 1.4(d)), and (2) improves its

overall power efficiency and the signal contrast at the correct output detector with respect to the detectors assigned to other classes (see Figs. 1.4(b), (c), (e), (f)). Therefore, for a given input illumination power and detector signal-to-noise ratio, the overall error rate of the all-optical network decreases as the number of diffractive layers increase. All these highlight the depth feature of a D^2NN.

This is not in contradiction with the fact that, for an all-optical D^2NN that is made of linear optical materials, the entire diffraction phenomenon that happens between the input and output planes can be squeezed into a single matrix operation (in reality, every material exhibits some volumetric and surface nonlinearities, and what we mean here by a linear optical material is that these effects are negligible). In fact, such an arbitrary mathematical operation defined by multiple learnable diffractive layers cannot be performed in general by a single diffractive layer placed between the same input and output planes; additional optical components/layers would be needed to all-optically perform an arbitrary mathematical operation that multiple learnable diffractive layers can in general perform. Our D^2NN framework creates a unique opportunity to use deep learning principles to design multiple diffractive layers, within a very tight layer-to-layer spacing of less than $50\times\lambda$, that collectively function as an all-optical classifier, and this framework will further benefit from nonlinear optical materials[15] and resonant optical structures to further enhance its inference performance.

In summary, the "depth" is a feature/property of a neural network, which means the network gets in general better at its inference and generalization performance with more layers. The mathematical origins of the depth feature for standard electronic neural networks relate to nonlinear activation function of the neurons. But this is not the case for a diffractive optical

network since it is a different type of a network, not following the same architecture or the same mathematical formalism of an electronic neural network.

Connectivity in diffractive optical networks

In a D^2NN design, the layer-to-layer connectivity of the optical network is controlled by several parameters: the axial distance between the layers (Δ_Z), the illumination wavelength (λ), the size of each fabricated neuron and the width of the diffractive layers. In our numerical simulations, we used a neuron size of approximately $0.53\times\lambda$. In addition, the height and width of each diffractive layer was set to include $200 \times 200 = 40K$ neurons per layer. In this arrangement, if the axial distance between the successive diffractive layers is set to be $\sim 40\times\lambda$ as in [15], then our D^2NN design becomes fully-connected. On the other hand, one can also design a much thinner and more compact diffractive network by reducing Δ_Z at the cost of limiting the connectivity between the diffractive layers. To evaluate the impact of this reduction in network connectivity on the inference performance of a diffractive neural network, we tested the performance of our D^2NN framework using $\Delta_Z = 4\times\lambda$, i.e., 10-fold thinner compared to our earlier discussed diffractive networks. With this partial connectivity between the diffractive layers, the blind testing accuracy for a 5-layer, phase-only D^2NN decreased from 97.18% ($\Delta_Z = 40\times\lambda$) to 94.12% ($\Delta_Z = 4\times\lambda$) for MNIST dataset (see Figs. 1.2(a) and 1.2(b), respectively). However, when the optical neural network with $\Delta_Z = 4\times\lambda$ was relaxed from phase-only modulation constraint to full complex modulation, the classification accuracy increased to 96.01% (Fig. 1.2(d)), partially compensating for the lack of full-connectivity. Similarly, for Fashion-MNIST dataset, the same compact architecture with $\Delta_Z = 4\times\lambda$ provided accuracy values of 85.98% and 88.54% for phase-only and complex-valued modulation schemes, as shown in Figs. 1.3(b) and 1.3(d), respectively, demonstrating the vital role of phase and amplitude modulation

capability for partially-connected, thinner and more compact optical networks.

Integration of diffractive neural networks with electronic networks: Performance analysis of D^{2}NN-based hybrid machine learning systems

Integration of passive diffractive neural networks with electronic neural networks (see e.g., Figs. 1.5(a) and 1.5(c)) creates some unique opportunities to achieve pervasive and low-power machine learning systems that can be realized using simple and compact imagers, composed of e.g., a few tens to hundreds of pixels per opto-electronic sensor frame. To investigate these opportunities, for both MNIST (Table 1.1) and Fashion-MNIST (Table 1.2) datasets, we combined our D^{2}NN framework (as an all-optical *front-end*, composed of 5 diffractive layers) with 5 different electronic neural networks considering various sensor resolution scenarios as depicted in Table 1.3. For the electronic neural networks that we considered in this analysis, in terms of complexity and the number of trainable parameters, a single fully-connected (FC) digital layer and a custom designed 4-layer convolutional neural network (CNN) (we refer to it as 2C2F-1 due to the use of 2 convolutional layers with a single feature and subsequent 2 FC layers) represent the lower end of the spectrum (see Tables 1.3-1.4); on the other hand, LeNet[25], ResNet-50[31] and another 4-layer CNN[33] (we refer to it as 2C2F-64 pointing to the use of 2 convolutional layers, subsequent 2 FC layers and 64 high-level features at its second convolutional layer) represent some of the well-established and proven deep neural networks

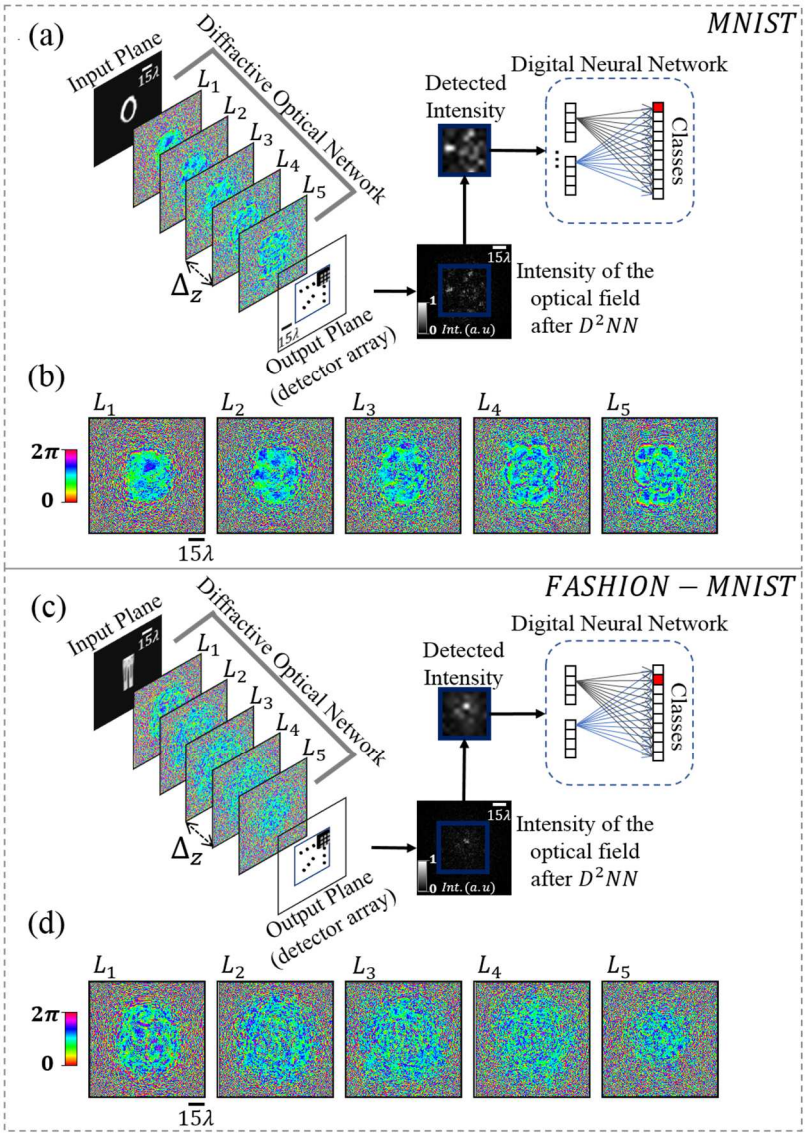

Fig. 1.5 D2NN-based hybrid neural networks. (a) The architecture of a hybrid (optical and electronic) classifier. (b) Final design of phase-only optical layers ($\boldsymbol{\Delta_z = 40}\times\lambda$) at the front-end of a hybrid handwritten digit classifier with a 10×10 opto-electronic detector array at the bridge/junction between the two modalities (optical vs. electronic). (c) and (d) are same as (a) and (b), except the latter are for Fashion-MNIST dataset. Input plane represents the plane of the input object or its data, which can also be generated by another optical imaging system or a lens, projecting an image of the object data onto this plane.

with more advanced architectures and considerably higher number of trainable parameters (see Table 1.3). All these digital networks used in our analysis, were individually placed after both a fully-connected ($\Delta_Z = 40\times\lambda$) and a partially-connected ($\Delta_Z = 4\times\lambda$) D^2NN design and the entire hybrid system in each case was *jointly* optimized at the second stage of the hybrid system training procedure detailed in the Methods section (see Fig. 1.6).

Among the all-optical D^2NN-based classifiers presented in the previous sections, the fully-connected ($\Delta_Z = 40\times\lambda$) complex modulation D^2NN designs have the highest classification accuracy values, while the partially-connected ($\Delta_Z = 4\times\lambda$) designs with phase-only restricted modulation are at the bottom of the performance curve (see the *all-optical* parts of Tables 1.1 and 1.2). Comparing the all-optical classification results based on a simple *max* operation at the output detector plane against the first rows of the "Hybrid Systems" sub-tables reported in Tables 1.1 and 1.2, we can conclude that the addition of a single FC layer (using 10 detectors), jointly-optimized with the optical part, can make up for some of the limitations of the D^2NN optical front-end design such as partial connectivity or restrictions on the neuron modulation function.

The 2nd, 3rd and 4th rows of the "Hybrid Systems" sub-tables reported in Tables 1.1 and 1.2 illustrate the classification performance of hybrid systems when the interface between the optical and electronic networks is a conventional focal plane array (such as a CCD or CMOS sensor array). The advantages of our D^2NN framework become more apparent for these cases, compared against traditional systems that have a conventional imaging optics-based front-end (e.g., a standard camera interface) followed by a digital neural network for which the classification accuracies are also provided at the bottom of Tables 1.1 and 1.2. From these comparisons reported in Tables 1.1 and 1.2, we can deduce that having a jointly-trained optical

and electronic network improves the inference performance of the overall system using low-end electronic neural networks as in the cases of a single FC network and 2C2F-1 network; also see Table 1.3 for a comparison of the digital neural networks employed in this work in terms of (1) the number of trainable parameters, (2) FLOPs, and (3) energy consumption. For example, when the 2C2F-1 network is used as the digital processing unit following a perfect imaging optics, the classification accuracies for MNIST (Fashion-MNIST) dataset are held as 89.73% (76.83%), 95.50% (81.76%) and 97.13% (87.11%) for 10×10, 25×25 and 50×50 detector arrays, respectively. However, when the same 2C2F-1 network architecture is enabled to jointly-evolve with e.g., the phase-only diffractive layers in a D^2NN front-end during the training phase, blind testing accuracies for MNIST (Fashion-MNIST) dataset significantly improve to 98.12% (89.55%), 97.83% (89.87%) and 98.50% (89.42%) for 10×10, 25×25 and 50×50 detector arrays, respectively. The classification performance improvement of the jointly-optimized hybrid system (diffractive + electronic network) over a perfect imager-based simple all-electronic neural network (e.g., 2C2F-1) is especially significant for 10×10 detectors (i.e., ~8.4% and ~12.7% for MNIST and Fashion-MNIST datasets, respectively). Similar performance gains are also achieved when single FC network is jointly-optimized with D^2NN instead of a perfect imaging optics/camera interface, preceding the all-electronic network as detailed in Tables 1.1 and 1.2. In fact, for some cases the classification performance of D^2NN-based hybrid systems, e.g. 5-layer, phase-only D^2NN followed by a single FC layer using any of the 10×10, 25×25 and 50×50 detectors arrays, shows a classification performance on par with a perfect imaging system that is followed by a more powerful, and energy demanding LeNet architecture (see Table 1.3).

Among the 3 different detector array arrangements that we investigated here, 10×10 detectors represent the case where the intensity on the opto-electronic sensor plane is severely

undersampled. Therefore, the case of 10×10 detectors represents a substantial loss of information for the imaging-based scenario (note that the original size of the objects in both image datasets is 28×28). This effect is especially apparent in Table 1.2, for Fashion-MNIST, which represents a more challenging dataset for object classification task, in comparison to MNIST. According to Table 1.2, for a computer vision system with a perfect camera interface and imaging optics preceding the opto-electronic sensor array, the degradation of the classification performance due to spatial undersampling varies between 3% to 5% depending on the choice of the electronic network. *However*, jointly-trained hybrid systems involving trainable diffractive layers maintain their classification performance even with ~7.8 times reduced number of input pixels (i.e., 10×10 pixels compared to the raw data, 28×28 pixels). For example, the combination of a fully-connected ($40\times\lambda$ layer-to-layer distance) D^2NN optical front-end with 5 phase-only (complex) diffractive layers followed by LeNet provides 90.24% (90.24%) classification accuracy for fashion products using a 10×10 detector array, which shows improvement compared to 87.44% accuracy that LeNet alone provides following a perfect imaging optics, camera interface. A similar trend is observed for all the jointly-optimized D^2NN-based hybrid systems, providing 3-5% better classification accuracy compared to the performance of all-electronic neural networks following a perfect imager interface with 10×10 detectors. Considering the importance of compact, thin and low-power designs, such D^2NN-based hybrid systems with significantly reduced number of opto-electronic pixels and an ultra-thin all-optical D^2NN front-end with a layer-to-layer distance of a few wavelengths cast a highly sought design to extend the applications of jointly-trained opto-electronic machine learning systems to various fields, without sacrificing their performance.

All-Optical

	$\Delta_z = 40\times\lambda$	$\Delta_z = 4\times\lambda$
Phase only	97.18	94.12
Complex	97.81	96.01

Hybrid Systems

# of detectors	Optical Modulation	Digital Neural Networks: Single FC Layer		2C2F-1		LeNet		2C2F-64		ResNet	
10	Phase only	97.65	93.12	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	Complex	98.02	95.96	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
10×10	Phase only	98.71	98.21	98.12	97.62	98.42	98.25	98.55	98.23	N/A	N/A
	Complex	98.29	98.20	98.35	97.60	98.59	98.25	98.56	98.31	N/A	N/A
25×25	Phase only	98.80	96.89	97.83	98.26	98.77	98.10	98.86	98.13	N/A	N/A
	Complex	98.64	97.50	98.37	98.14	98.62	98.10	98.57	98.18	N/A	N/A
50×50	Phase only	98.82	98.07	98.50	97.88	98.65	97.93	98.92	98.35	98.97	98.09
	Complex	98.81	97.99	98.17	98.22	98.56	98.06	98.63	98.32	98.54	98.11

Imaging Optics Based Classification Systems

# of detectors	Single FC Layer	2C2F-1	LeNet	2C2F-64	ResNet
10×10	91.50	89.73	98.36	98.18	N/A
25×25	92.91	95.50	98.83	98.99	N/A
50×50	92.44	97.13	98.95	99.04	99.53

Table 1.1 Blind testing accuracies (reported in percentage) for all-optical (D2NN only), D2NN and perfect imager-based hybrid systems used in this work for MNIST dataset.

On the other hand, for designs that involve higher pixel counts and more advanced electronic neural networks (with higher energy and memory demand), our results reveal that D^2NN based hybrid systems perform worse compared to the inference performance of perfect imager-based computer vision systems. For example, based on Tables 1.1 and 1.2 one can infer that using ResNet as the electronic neural network of the hybrid system with 50x50 pixels, the discrepancy between the two approaches (D^2NN vs. perfect imager based front-end choices) is ~0.5% and ~4% for MNIST and Fashion-MNIST datasets, respectively, in favor of the perfect imager front-end. We believe this inferior performance of the jointly-optimized D^2NN-based

hybrid system (when higher pixel counts and more advanced electronic networks are utilized) is related to sub-optimal convergence of the diffractive layers in the presence of a powerful electronic neural network that is by and large determining the overall loss of the jointly-optimized hybrid network during the training phase. In other words, considering the lack of non-linear activation functions within the D^2NN layers, a powerful electronic neural network at the back-end hinders the evolution of the optical front-end during training phase due to its relatively superior approximation capability. Some of the recent efforts in the literature to provide a better understanding of the inner workings of convolutional neural networks[34,35] might help us to devise more efficient learning schemes to overcome this "shadowing" behavior in order to improve the inference performance of our jointly-optimized D^2NN-based hybrid systems. Extending the fundamental design principles and methods behind diffractive optical networks to operate under spatially and/or temporally incoherent illumination is another intriguing research direction stimulated by this work, as most computer vision systems of today rely on incoherent ambient light conditions. Finally, the flexibility of the D^2NN framework paves the way for broadening our design space in the future to metasurfaces and metamaterials through essential modifications in the parameterization of the optical modulation functions [36,37].

1.3 Methods

Diffractive network architectures

In our diffractive neural network model, the input plane represents the plane of the input object or its data, which can also be generated by another optical imaging system or a lens, e.g., by projecting an image of the object data. Input objects were encoded in amplitude channel (MNIST) or phase channel (Fashion-MNIST) of the input plane and were illuminated with a

uniform plane wave at a wavelength of λ to match the conditions introduced in [15] for all-optical classification. In the hybrid system simulations presented in Tables 1.1 and 1.2, on the other hand, the objects in both datasets were represented as amplitude objects at the input plane, providing a fair comparison between the two tables.

Optical fields at each plane of a diffractive network were sampled on a grid with a spacing of ~0.53λ in both x and y directions. Between two diffractive layers, the free-space propagation was calculated using the angular spectrum method[15]. Each diffractive layer, with a neuron size of $0.53\lambda \times 0.53\lambda$, modulated the incident light in phase and/or amplitude, where the modulation value was a trainable parameter and the modulation method (phase-only or complex) was a pre-defined design parameter of the network. The number of layers and the axial distance from the input plane to the first diffractive layer, between the successive diffractive layers, and from the last diffractive layer to the detector plane were also pre-defined design parameters of each network. At the detector plane, the output field intensity was calculated.

Forward optical model and training loss functions

The physical model in our diffractive framework does not rely on small diffraction angles or the Fresnel approximation and is not restricted to far-field analysis (Fraunhofer diffraction) [38,39]. Following the Rayleigh-Sommerfeld equation, a single neuron can be considered as the secondary source of wave $w_i^l(x, y, z)$, which is given by:

$$w_i^l(x, y, z) = \frac{z - z_i}{r^2}\left(\frac{1}{2\pi r} + \frac{1}{j\lambda}\right)\exp\left(\frac{j2\pi r}{\lambda}\right) \tag{1.3}$$

where $r = \sqrt{(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2}$ and $j = \sqrt{-1}$. Treating the input plane as the 0th

layer, then for l^{th} layer ($l \geq 1$), the output field can be modeled as:

$$u_i^l(x,y,z) = w_i^l(x,y,z) \cdot t_i^l(x_i,y_i,z_i) \cdot \sum_k u_k^{l-1}(x_i,y_i,z_i)$$

$$= w_i^l(x,y,z) \cdot |A| \cdot e^{j\Delta\theta}, \quad (1.4)$$

where $u_i^l(x,y,z)$ denotes the output of the i^{th} neuron on l^{th} layer located at (x,y,z), the t_i^l denotes the complex modulation, i.e., $t_i^l(x_i,y_i,z_i) = a_i^l(x_i,y_i,z_i)exp(j\phi_i^l(x_i,y_i,z_i))$. In eq. (1.4), $|A|$ is the relative amplitude of the secondary wave, and $\Delta\theta$ refers to the additional phase delay due to the input wave at each neuron, $\sum_k u_k^{l-1}(x_i,y_i,z_i)$, and the complex-valued neuron modulation function, $t_i^l(x_i,y_i,z_i)$.

To perform classification by means of all-optical diffractive networks with minimal post-processing (i.e., using only a max operation), we placed discrete detectors at the output plane. The number of detectors (D) is equal to the number of classes in the target dataset. The geometrical shape, location and size of these detectors ($6.4\lambda \times 6.4\lambda$) were determined before each training session. Having set the detectors at the output plane, the final loss value (L) of the diffractive neural network is defined through two different loss functions and their impact on D^2NN based classifiers were explored (see the *Results* section). The first loss function was defined using the mean squared error (MSE) between the output plane intensity, S^{l+1}, and the target intensity distribution for the corresponding label, G^{l+1}, i.e.,

$$L = \frac{1}{K}\sum_i^K \left(S_i{}^{l+1} - G_i{}^{l+1}\right)^2, \quad (1.5)$$

where K refers to the total number of sampling points representing the entire diffraction pattern at the output plane.

The second loss function used in combination with our all-optical D^{2}NN framework is the cross-entropy. To use the cross-entropy loss function, an additional *softmax* layer is introduced and applied on the detected intensities (only during the training phase of a diffractive neural network design). Since *softmax* function is *not* scale invariant[40], the measured intensities by D detectors at the output plane are normalized such that they lie in the interval (0,10) for each sample. With I_l denoting the total optical signal impinging onto the l^{th} detector at the output plane, the normalized intensities, I_l', can be found by,

$$I_l' = \frac{I_l}{\max\{I_l\}} \times 10. \tag{1.6}$$

In parallel, the cross-entropy loss function can be written as follows:

$$L = -\sum_l^D g_l \log(p_l), \tag{1.7}$$

where $p_l = \frac{e^{I_l'}}{\sum_l^D e^{I_l'}}$ and g_l refer to the l^{th} element in the output of the *softmax* layer, and the l^{th} element of the ground truth label vector, respectively.

A key difference between the two loss functions is already apparent from eq. (1.5) and eq. (1.7). While the MSE loss function is acting on the entire diffraction signal at the output plane of the diffractive network, the *softmax-cross-entropy* is applied to the detected optical signal values ignoring the optical field distribution outside of the detectors (one detector is assigned per class). This approach based on *softmax-cross-entropy* loss brings additional degrees-of-freedom to the diffractive neural network training process, boosting the final classification performance as discussed in the *Results* section, at the cost of reduced diffraction efficiency and signal contrast at the output plane.

For both the imaging optics-based and hybrid (D^{2}NN + electronic) classification systems

presented in Tables 1.1 and 1.2, the loss functions were also based on *softmax-cross-entropy*.

Diffractive network training

All neural networks (optical and/or digital) were simulated using Python (v3.6.5) and TensorFlow (v1.10.0, Google Inc.) framework. All-optical, hybrid and electronic networks were trained for 50 epochs using a desktop computer with a GeForce GTX 1080 Ti Graphical Processing Unit, GPU and Intel(R) Core (TM) i9-7900X CPU @3.30GHz and 64GB of RAM, running Windows 10 operating system (Microsoft).

Two datasets were used in the training of the presented classifiers: MNIST and Fashion-MNIST. Both datasets have 70,000 objects/images, out of which we selected 55,000 and 5,000 as training and validation sets, respectively. Remaining 10,000 were reserved as the test set. During the training phase, after each epoch we tested the performance of the current model in hand on the 5K validation set and upon completion of the 50^{th} epoch, the model with the best performance on 5K validation set was selected as the final design of the network models. All the numbers reported in this work are blind testing accuracy results held by applying these selected models on the 10K test sets.

The trainable parameters in a diffractive neural network are the modulation values of each layer, which were optimized using a back-propagation method by applying the adaptive moment estimation optimizer (Adam)[41] with a learning rate of 10^{-3}. We chose a diffractive layer size of 200×200 neurons per layer, which were initialized with π for phase values and 1 for amplitude values. The training time was approximately 5 hours for a 5-layer D^2NN design with the hardware outlined above.

D^2NN-based hybrid network training

To further explore the potentials of D^2NN framework, we co-trained diffractive network layers together with digital neural networks to form hybrid systems. In these systems, the detected intensity distributions at the output plane of the diffractive network were taken as the input for the digital neural network at the back-end of the system.

To begin with, keeping the optical architecture and the detector arrangement at the output plane of the diffractive network same as in the all-optical case, a single fully-connected layer was introduced as an additional component (replacing the simplest *max* operations in an all-optical network), which maps the optical signal values coming from *D* individual detectors into a vector of the same size (i.e., the number of classes in the dataset). Since there are 10 classes in both MNIST and Fashion-MNIST datasets, this simple fully-connected digital structure brings additional 110 trainable variables (i.e., 100 coefficients in the weight matrix and 10 bias terms) into our hybrid system.

We have also assessed hybrid configurations that pair D^2NNs with CNNs, a more popular architecture than fully-connected networks for object classification tasks. In such an arrangement, when the optical and electronic parts are directly cascaded and jointly-trained, the inference performance of the overall hybrid system was observed to stagnate at a local minimum. As a possible solution to this issue, we divided the training of the hybrid systems into two stages as shown in Fig. 1.6. In the first stage, the detector array was placed right after the D^2NN optical front-end, which was followed by an additional, virtual optical layer, acting as an all-optical classifier (see Fig. 1.6(a)). We emphasize that this additional optical layer *is not* part of the hybrid system at the end; instead it will be replaced by a digital neural network in the second

stage of our training process. The sole purpose of two-stage training arrangement used for hybrid systems is to find a better initial condition for the D^2NN that precedes the detector array, which is the interface between the fully optical and electronic networks.

In the second stage of our training process, the already trained 5-layer D^2NN optical front-end (preceding the detector array) was cascaded and jointly-trained with a digital neural network. It is important to note that the digital neural network in this configuration was trained from scratch. This type of procedure "resembles" transfer learning, where the additional layers (and data) are used to augment the capabilities of a trained model[42].

Using the above described training strategy, we studied the impact of different configurations, by increasing the number of detectors forming an opto-electronic detector array, with a size of 10×10, 25×25 and 50×50 pixels. Having different pixel sizes (see Table 1.3), all the three configurations (10×10, 25×25 and 50×50 pixels) cover the central region of approximately $53.3\,\lambda \times 53.3\,\lambda$ at the output plane of the D^2NN. Note that each detector configuration represents different levels of spatial undersampling applied at the output plane of a D^2NN, with 10×10 pixels corresponding to the most severe case. For each detector configuration, the first stage of the hybrid system training, shown in Fig. 1.6(a) as part of Appendix A, was carried out for 50 epochs providing the initial condition for 5-layer D^2NN design before the joint-optimization phase at the second stage. These different initial optical front-end designs along with their corresponding detector configurations were then combined and jointly-trained with various digital neural network architectures, simulating different hybrid systems (see Fig. 1.6(b) and Fig 1.5). At the interface of optical and electronic networks, we introduced a batch normalization layer applied on the detected intensity distributions at the sensor.

For the digital part, we focused on five different networks representing different levels complexity regarding (1) the number of trainable parameters, (2) the number of FLOPs in the forward model and (3) the energy consumption; see Table 1.3. This comparative analysis depicted in Table 1.3 on energy consumption assumes that 1.5pJ is needed for each multiply-accumulate (MAC)[43] and based on this assumption, the 4th column of Table 1.3 reports the energy needed for each network configuration to classify an input image. The first one of these digital neural networks was selected as a single fully-connected (FC) network connecting every pixel of detector array with each one of the 10 output classes, providing as few as 1,000 trainable parameters (see Table 1.3 for details). We also used the 2C2F-1 network as a custom designed CNN with 2 convolutional and 2 FC layers with only a single filter/feature at each convolutional layer (see Table 1.4). As our 3rd network, we used LeNet[25] which requires a certain input size of 32×32 pixels, thus the detector array values were resized using bilinear interpolation before being fed into the electronic neural network. The fourth network architecture that we used in our comparative analysis (i.e., 2C2F-64), as described in [33], has 2 convolutional and 2 fully-connected layers similar to the second network, but with 32 and 64 features at the first and second convolutional layers, respectively, and has larger FC layers compared to the 2C2F-1 network. Our last network choice was ResNet-50[31] with 50 layers, which was only jointly-trained using the 50×50 pixel detector configuration, the output of which was resized using bilinear interpolation to 224×224 pixels before being fed into the network. The loss function of the D^2NN-based hybrid system was calculated by cross-entropy, evaluated at the output of the digital neural network.

As in D^2NN-based hybrid systems, the objects were assumed to be purely amplitude modulating functions for perfect imager-based classification systems presented in Tables 1.1 and

1.2; moreover, the imaging optics or the camera system preceding the detector array is assumed to be diffraction limited which implies that the resolution of the captured intensity at the detector plane is directly limited by the pixel pitch of the detector array. The digital network architectures and training schemes were kept identical to D^2NN-based hybrid systems to provide a fair comparison. Also, worth noting, no data augmentation techniques have been used for any of the networks presented in this manuscript.

Details of D^2NN-based hybrid network training procedure

We introduced a two-stage training pipeline for D^2NN-based hybrid classifiers as mentioned in the previous sub-section. The main reason behind the development of this two-stage training procedure stems from the unbalanced nature of the D^2NN-based hybrid systems, especially if the electronic part of the hybrid system is a powerful deep convolutional neural network (CNN) such as ResNet. Being the more powerful of the two and the latter in the information processing order, deep CNNs adapt and converge faster than D^2NN-based optical front-ends. Therefore, directly cascading and jointly-training D^2NNs with deep CNNs offer a suboptimal solution on the classification accuracy of the overall hybrid system.

Figure 1.6 illustrates the two-step training procedure for D^2NN-based hybrid system training, which was used for the results reported in Tables 1.1 and 1.2. In the first step, we introduce the detector array model that is going to be the interface between the optical and the electronic networks. An additional virtual diffractive layer is placed right after the detector plane (see Fig. 1.6(a)). We model the detector array as an intensity sensor (discarding the phase information). Implementing such a detector array model with an *average pooling* layer which has *strides* as large as its kernel size on both directions, the detected intensity, I_A, is held at the focal

plane array. In our simulations, the size of I_A was 10×10, 25×25 or 50×50, depending on the choice of the detector array used in our design. To further propagate this information through the virtual 1-Layer optical classifier (Fig. 1.6(a)), I_A is interpolated using the *nearest neighbour* method back to the object size at the input plane. Denoting this interpolated intensity as $I_A^{'}$, the propagated field is given by $\sqrt{I_A^{'}}$ (see Fig. 1.6(a)). It is important to note that the phase information at the output plane of the D^2NN preceding the detector array is entirely discarded, thus the virtual classifier decides solely based on the measured intensity (or underlying amplitude) as it would be the case for an electronic network.

After training this model for 50 epochs, the layers of the diffractive network preceding the detector array are taken as the ***initial*** condition for the optical part in the second stage of our training process (see Fig. 1.6(b)). Starting from the parameters of these diffractive layers, the second stage of our training simply involves the ***simultaneous*** training of a D^2NN-based optical part and an electronic network at the back-end of the detector array bridging two modalities as shown in Fig. 1.6(b). In this second part of the training, the detector array model is kept identical with the first part and the electronic neural network is trained from scratch with optical and electronic parts having equal learning rates (10^{-3}).

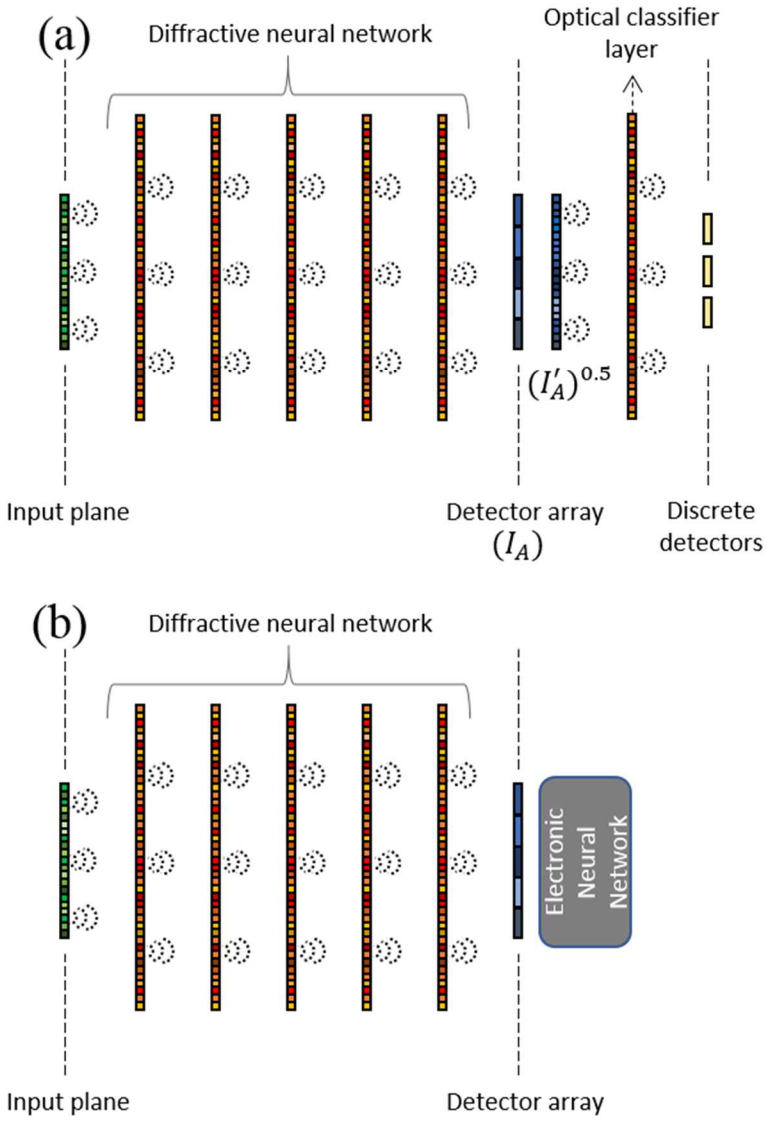

Fig. 1.6 Hybrid system training procedure. (a) The first stage of the hybrid system training. (b) The second stage of the hybrid system training starts with the already trained diffractive layers (first 5 layers) from part (a) and an electronic neural network, replacing the operations after intensity detection at the sensor. Note that the spherical waves between the consequent layers in (a) and (b) illustrate free space wave propagation.

All-Optical

	$\Delta_z = 40\times\lambda$	$\Delta_z = 4\times\lambda$
Phase only	88.57	85.69
Complex	88.94	88.29

Hybrid Systems

		Digital Neural Networks									
# of detectors	Optical Modulation	Single FC Layer		2C2F-1		LeNet		2C2F-64		ResNet	
10	Phase only	88.88	87.76	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
	Complex	89.57	88.40	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
10×10	Phase only	90.04	88.84	89.55	88.83	90.24	89.19	90.08	89.76	N/A	N/A
	Complex	89.96	88.88	89.26	89.43	90.24	89.55	89.92	89.88	N/A	N/A
25×25	Phase only	90.08	88.75	89.87	89.02	89.96	89.20	89.84	89.66	N/A	N/A
	Complex	90.25	88.57	89.94	89.50	89.79	89.64	89.75	89.83	N/A	N/A
50×50	Phase only	90.22	89.43	89.42	89.72	89.71	89.24	89.66	90.30	89.20	89.43
	Complex	89.54	89.45	90.11	89.79	89.74	89.76	89.29	90.45	89.29	89.40

Imaging Optics Based Classification Systems

# of detectors	Single FC Layer	2C2F-1	LeNet	2C2F-64	ResNet
10×10	81.20	76.83	87.44	88.11	N/A
25×25	84.47	81.76	90.19	91.6	N/A
50×50	84.49	87.11	90.33	91.9	93.46

Table 1.2 Blind testing accuracies (reported in percentage) for all-optical (D2NN only), D2NN and perfect imager-based hybrid systems used in this work for Fashion-MNIST dataset.

Digital Neural Networks	Trainable Parameters	FLOPs	Energy Consumption (J/image)	Detector Configuration
Single FC Layer	1000	2000	1.5×10^{-9}	10×10
	6250	12500	9.5×10^{-9}	25×25
	25000	50000	3.8×10^{-8}	50×50
2C2F-1	615	3102	2.4×10^{-9}	10×10
	825	9048	7.0×10^{-9}	25×25
	3345	43248	3.3×10^{-8}	50×50
LeNet[25]	60840	1×10^{6}	7.5×10^{-7}	10×10
				25×25
				50×50
2C2F-64[33]	3.3×10^{5}	3.1×10^{6}	2.4×10^{-6}	10×10
	2.4×10^{6}	2.5×10^{7}	1.9×10^{-5}	25×25
	9.5×10^{6}	8.7×10^{7}	6.5×10^{-5}	50×50
ResNet[31]	25.5×10^{6}	4×10^{9}	3×10^{-3}	50×50

Table 1.3 Comparison of electronic neural networks in terms of the number of trainable parameters

	Network architecture							
Layer Type	Conv layer 1			Conv layer 2			FC layer 1	FC layer 2
Activation	ReLU			ReLU			ReLU	Softmax
Detector configuration	kernel	Feature map	Stride	kernel	Feature map	Stride	Number of neurons	Number of neurons
10×10	6×6	1	1	3×3	1	1	30	10
25×25			2			2		
50×50			2			2		

Table 1.4 Parameters of the custom designed network architecture which we refer to as 2C2F-1.

Chapter 2 Misalignment Resilient Diffractive Optical Networks

Parts of this chapter have previously been published in D. Mengu et al. "Misalignment Resilient Diffractive Optical Networks" Nanophotonics, vol. 9, no. 13, 2020, pp. 4207-4219 (2020), DOI: 10.1515/nanoph-2020-0291. In this chapter, the D^2NN framework is extended to mitigate the impact of physical error sources in the forward optical model of a fabricated diffractive network.

As an optical machine learning framework, D^2NN takes advantage of data-driven training methods used in deep learning to devise light-matter interaction in 3D for performing a desired statistical inference task. Multi-layer optical object recognition platforms designed with this diffractive framework have been shown to generalize to unseen image data achieving e.g., >98% blind inference accuracy for hand-written digit classification. The multi-layer structure of diffractive networks offers significant advantages in terms of their diffraction efficiency, inference capability and optical signal contrast. However, the use of multiple diffractive layers also brings practical challenges for the fabrication and alignment of these diffractive systems for accurate optical inference. Here, we introduce and experimentally demonstrate a new training scheme that significantly increases the robustness of diffractive networks against 3D misalignments and fabrication tolerances in the physical implementation of a trained diffractive network. By modeling the undesired layer-to-layer misalignments in 3D as continuous random variables in the optical forward model, diffractive networks are trained to maintain their inference accuracy over a large range of misalignments; we term this diffractive network design as *vaccinated* D^2NN (v-D^2NN). We further extend this vaccination strategy to the training of diffractive networks that use differential detectors at the output plane as well as to jointly-trained

hybrid (optical-electronic) networks to reveal that all of these diffractive designs improve their resilience to misalignments by taking into account possible 3D fabrication variations and displacements during their training phase.

2.1 Introduction

Deep learning has been redefining the state-of-the-art for processing various signals collected and digitized by different sensors, monitoring physical processes for e.g., biomedical image analysis[44–47], speech recognition[48,49] and holography[50–53], among many others[54–60]. Furthermore, deep learning and related optimization tools have been harnessed to find data-driven solutions for various inverse problems arising in, e.g., microscopy[61–65], nanophotonic designs and plasmonics[66–68]. These demonstrations and others have been motivating some of the recent advances in optical neural networks and related optical computing techniques that aim to exploit the computational speed, power-efficiency, scalability and parallelization capabilities of optics for machine intelligence applications[20,22,69–86].

Toward this broad goal, Diffractive Deep Neural Networks (D^2NN)[77–80] have been introduced as a machine learning framework that unifies deep learning-based training of matter with the physical models governing light propagation to enable all-optical inference through a set of diffractive layers. The training stage of a diffractive network is performed using a computer, and relies on deep learning and error backpropagation methods to tailor the light-matter interaction across a set of diffractive layers that collectively perform a given machine learning task, e.g., object classification. Previous studies on D^2NNs have demonstrated the generalization capability of these multi-layer diffractive network designs to new, unseen image data. For example, using a 5-layer diffractive network architecture, >98% and >90% all-optical blind

testing accuracies have been reported [79] for the classification of the images of handwritten digits (MNIST) [87] and fashion products (Fashion-MNIST) [88] that are encoded in the amplitude and phase channels of the input plane, respectively. Successful experimental demonstrations of these all-optical classification systems have been reported using 3D-printed diffractive layers that conduct inference by modulating the incoming object wave at terahertz (THz) wavelengths.

Despite the lack of nonlinear optical elements in these previous implementations, diffractive optical networks have been shown to offer significant advantages in terms of (1) inference accuracy, (2) diffraction efficiency and (3) signal contrast, when the number of successive diffractive layers in the network design is increased [78]. A similar depth advantage was also demonstrated in [80], where instead of a statistical inference task such as image classification, the D^2NN framework was utilized to solve an inverse design problem to achieve e.g., spatially-controlled wavelength de-multiplexing of a broadband source. While these multi-layer diffractive architectures offer significantly better performance for generalization and application-specific design merits, they also pose practical challenges for the fabrication and opto-mechanical assembly of these trained diffractive models.

Here, we present a training scheme that substantially increases the robustness of diffractive optical networks against physical misalignments and fabrication tolerances. Our scheme models and introduces these undesired system variations and layer-to-layer misalignments as continuous random variables during the deep learning-based training of the diffractive model to significantly improve the error tolerance margins of diffractive optical networks; this process of introducing random misalignments during the training phase will be termed as *vaccination* of the diffractive network, and the resulting designs will be referred to as

vaccinated D^2NNs (v-D^2NNs). To demonstrate the efficacy of our strategy, we trained diffractive network models composed of 5 diffractive layers for all-optical classification of handwritten digits, where we utilized in the training phase independent and uniformly distributed displacement/misalignment vectors for x, y, and z directions of each diffractive layer. Our results indicate that v-D^2NN framework enables the design of diffractive optical networks that can maintain their object recognition performance against severe layer-to-layer misalignments, providing nearly flat blind inference accuracies within the displacement/misalignment range adopted in the training.

To experimentally demonstrate the success of v-D^2NN framework we also compared two 3D-printed diffractive networks, each with 5 diffractive layers that were designed for hand-written digit classification under monochromatic THz illumination (λ = ~0.75 mm): the first network model was designed without the presence of any misalignments (non-vaccinated) and the second one was designed as a v-D^2NN. After the fabrication of each diffractive network, the 3rd diffractive layer was on purpose misaligned to different 3D positions around its ideal location. The experimental results confirmed our numerical analysis to reveal that the v-D^2NN design can preserve its inference accuracy despite a wide range of physical misalignments, while the standard D^2NN design frequently failed to recognize the correct data class due to these purposely-introduced misalignments.

We also combined our v-D^2NN framework with the differential diffractive optical networks [79] and the jointly-trained optical-electronic (hybrid) neural network systems. Differential diffractive classification systems assign a pair of detectors (generating one positive and one negative signal) for each data class to mitigate the strict non-negativity constraint of

optical intensity, and were demonstrated to offer superior inference accuracy compared to standard diffractive designs [79]. When trained against misalignments using the presented v-D^2NN framework, differential diffractive networks are also shown to preserve their performance advantages for all-optical classification. However, both differential and standard diffractive networks fall short in matching the adaptation capabilities of a hybrid diffractive network system that uses a modest, single-layer fully-connected architecture with only 110 learnable parameters in the electronic domain, following the diffractive optical front-end.

In addition to misalignment related errors, the presented vaccination framework can also be adopted to mitigate other error sources in diffractive network models, e.g., detection noise and fabrication imperfections or artefacts, provided that the approximate analytical models and the probability distributions of these factors are utilized during the training stage. We anticipate that v-D^2NNs will be the gateway of diffractive optical networks and the related hybrid neural network schemes towards practical machine vision and sensing applications, by mitigating various sources of error between the training forward models and the corresponding physical hardware implementations. Furthermore, the presented methodology of designing misalignment and noise resilient physical machine learning models can be broadly applicable to other optical learning platforms, regardless of their physical dimensions and selected operation wavelengths.

2.2 Results

Figure 2.1 illustrates three different types of diffractive optical network-based object recognition systems investigated in this work. We focused on 5-layer diffractive optical network architectures as shown in Fig. 2.1 that are fully-connected, meaning that the half cone angle of the secondary wave created by the diffractive features (neurons) of size, e.g., $\delta=0.53\lambda$, is large

enough to enable communication between all the features on two successive diffractive layers that are placed e.g., 40λ apart in axial direction. On the transverse plane, each diffractive layer extends from -100×δ to 100×δ on x and y directions around the optical axis, and therefore the edge length of each diffractive surface in total is 200×δ (~106.66λ). With this outlined diffractive network architecture, the standard D^2NN training routine updates the trainable parameters of the diffractive layers at every iteration based on the mean gradient computed over a batch of training samples with respect to a loss function, specifically tailored for the desired optical machine learning application, e.g., cross-entropy for supervised object recognition systems [78], until a convergence criterion is satisfied. Since this conventional training approach assumes perfect alignment throughout the training, the sources of statistical variations in the resulting model are limited to the initial condition of the diffractive network parameter space and the sequence of the training data introduced to the network.

Training and testing of v-D2NNs

The training of vaccinated diffractive optical networks mainly follows the same steps as the standard D^2NN framework; except, it additionally incorporates system errors, e.g. misalignments, based on their probability distribution functions into the optical forward model. In this work, we modelled each orthogonal component of the undesired 3D displacement vector of each diffractive layer, $D = (D_X , D_y , D_Z)$, as uniformly distributed, independent random variables as follows;

$$D_X \sim U(-\Delta_X, \Delta_X),$$

$$D_y \sim U(-\Delta_y, \Delta_y),$$

$$D_Z \sim U(-\Delta_Z, \Delta_Z), \tag{2.8}$$

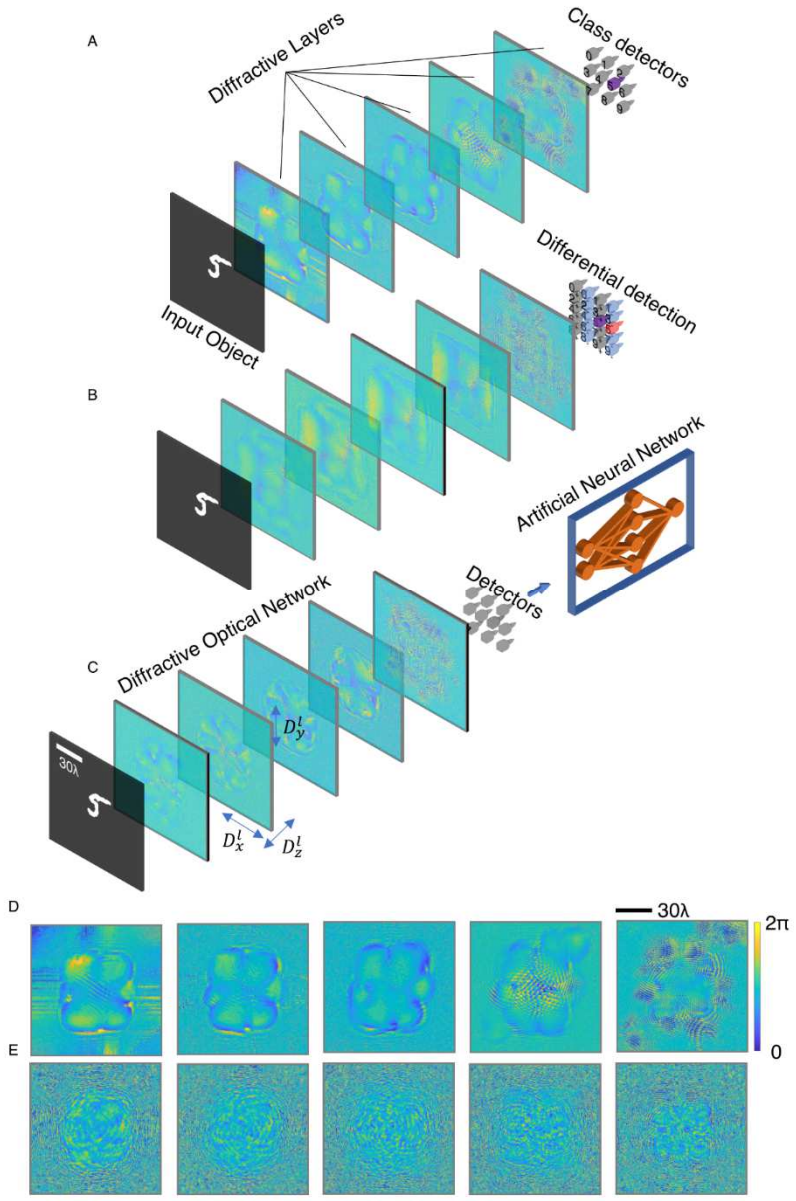

Fig. 2.1 Different types of D2NN-based image classification systems. A Standard D^2NN framework trained for all-optical classification of handwritten digits. Each detector at the output plane represents a data class. B Differential D^2NN trained for all-optical classification of handwritten digits. Each data class is represented by a pair

of detectors at the output plane, where the normalized difference between these detector pairs represents the class scores. C Jointly-trained hybrid (optical-electronic) network system trained for classification of handwritten digits. The optical signals collected at the output detectors are used as inputs to the electronic neural network at the back-end, which is used to output the final class scores. D Phase profiles computed by the deep learning-based training for a 5-layer diffractive optical network that is vaccinated against both lateral and axial misalignments for the task of handwritten digit classification. The layers of this diffractive network were fabricated using 3D printing as shown in Fig. 5D and experimentally tested using the setup shown in Fig. 5E. E Same as D, except the diffractive network represents a non-vaccinated, error-free design.

where $\Delta*$ denotes the shift along the corresponding axis, (*), reflecting the uncertainty in our physical assembly/fabrication of the diffractive model. During the training, the random displacement vector of *each* diffractive layer, **D**, takes different values sampled from the probability distribution of its components, D_X, D_Y and D_Z, for each batch of training samples. Consequently, the location of layer *l* at i^{th} iteration/batch, $\mathbf{L^{(l,i)}}$, can be expressed as;

$$L^{(l,i)} = (L_x^l, L_y^l, L_z^l) + (D_x^{(l,i)}, D_y^{(l,i)}, D_z^{(l,i)}), \tag{2.2}$$

where the first and the second vectors on the right-hand side denote the ideal location of the diffractive layer *l*, and a random realization of the displacement vector, $\mathbf{D^{(l,i)}}$, of layer *l* at the training iteration *i*, respectively. The displacement vector of each layer is independently determined, i.e., each layer of a diffractive network model can move within the displacement ranges depicted in Eq. (2.1) without any dependence on the locations of the other diffractive layers.

Opto-mechanical assembly and fabrication systems, in general, use different mechanisms to control the lateral and axial positioning of optical components. Therefore, we split our numerical investigation of the vaccination process into two: the lateral and axial misalignment cases. For the vaccination of diffractive optical network models against layer-to-layer misalignments on the transverse plane, we assumed D_X and D_Y are i.i.d random variables during the training, i.e. they are independent with a parameter of $\Delta_X = \Delta_Y = \Delta_{tr}$, and D_Z was set to be 0. The axial case, on the other hand, sets Δ_{tr} to be 0 throughout the training leaving $D_Z \sim U(-\Delta_{(z,tr)}, \Delta_{(z,tr)})$ as the only source of inter-layer misalignments.

Following a similar path with the training, the blind testing of the presented diffractive network models updates the random displacement vector of each layer l, $\mathbf{D^{(l,m)}}$, for *each* test sample m. The reported accuracies throughout our analyses reflects the *blind testing accuracies* computed over the 10K image test set of MNIST digits where *each test sample propagates through a diffractive network model that experiences a different realization of the random variables* depicted in Eq. (2.1) for each diffractive layer, i.e. *there are 10K different configurations that a diffractive network model was misaligned throughout the testing stage*. Furthermore, similar to the training process, during the blind testing against lateral misalignments, it was assumed that D_X and D_Y are i.i.d random variables with $\Delta_X = \Delta_Y = \Delta_{test}$, and similarly, the axial displacements or misalignments were determined by $D_Z \sim U(-\Delta_{(z,test)}, \Delta_{(z,test)})$.

Misalignment analysis of all-optical and hybrid diffractive systems

Figures 2.2A and 2.2D illustrate the blind testing accuracies provided by the standard diffractive optical network architecture (Fig. 2.1A) trained against various levels of undesired axial and lateral misalignments, respectively. Focusing on the testing accuracy curve obtained by the error-free design (dark blue) in Figs. 2.2A and 2.2D, it can be noticed that the diffractive optical networks are more susceptible to lateral misalignments compared to axial misalignments. For instance, when Δ_{test} is taken as 2.12λ, inducing random lateral fluctuations on each diffractive layer's location around the optical axis, the blind testing accuracy achieved by the non-vaccinated standard diffractive optical network decreases to 38.40% from 97.77% (obtained in the absence of misalignments). As we further increase the level of lateral misalignments, the error-free diffractive optical network almost completely loses its inference capability by achieving, e.g. 19.24% blind inference accuracy for $\Delta_{test}=4.24\lambda$ (i.e., the misalignment range in each lateral direction of a diffractive layer is -8δ to 8δ). On the other hand, when the diffractive layers are randomly misaligned on the longitudinal direction alone, the inference performance does not drop as excessively as the lateral misalignment case; for example, even when $\Delta_{(z,test)}$ becomes as large as 19.2λ, the error-free diffractive network manages to obtain an inference accuracy of 49.8%.

As demonstrated in Fig. 2.2D, the rapid drop in the testing accuracy of diffractive optical classification systems under physical misalignments can be mitigated by using the v-D^2NN framework. Since v-D^2NN training introduces displacement errors in the training stage, the diffractive optical networks can adopt to those variations preserving their inference performance over large misalignment margins. As an example, the 38.40% blind testing accuracy achieved by the non-vaccinated diffractive design with a lateral misalignment range of $\Delta_{test}=2.12\lambda$, can be increased to 94.44% when the same architecture is trained with a similar error range using the

presented vaccination framework (see the purple line in Fig. 2.2D). On top of that, the vaccinated design does not compromise the performance of the all-optical object recognition systems when the ideal conditions are satisfied. Compared to the 97.77% accuracy provided by the error-free design, this new vaccinated network (purple line in Fig. 2.2D) obtains 96.1% in the absence of misalignments. In other words, the ~56% inference performance gain of the vaccinated diffractive network under physical misalignments comes at the expense of only 1.67% accuracy loss when the opto-mechanical assembly perfectly matches the numerical training model. In case the level of misalignment-related imperfections in the fabrication of the diffractive network is expected to be even smaller, one can design improved v-D^{2}NN models that achieve e.g., 97.38%, which corresponds to only 0.39% inference accuracy loss compared to the error-free models at their peak (perfect alignment case) while at the same time providing >4% blind testing accuracy improvement under mild misalignment, i.e., $\Delta_{test}=0.53\lambda$. Similarly, when we compare the blind inference curves of the error-free and vaccinated network designs in Fig. 2.2A, one can notice that the v-D^{2}NN framework can easily recover the performance of the diffractive digit classification networks in the case where the displacement errors are restricted to be on the longitudinal axis. For example, with $\Delta_{(z,test)}=2.4\lambda$, the inference accuracy of the error-free diffractive network (dark blue) is reduced to 94.88%, while a vaccinated diffractive network that was already trained against the same level of misalignment, $\Delta_{(z,tr)}=2.4\lambda$ (yellow), retains 97.39% blind inference accuracy under the same level of axial misalignment.

Next, we combined our v-D^{2}NN framework with the *differential* diffractive network architecture: the blind testing results of various differential handwritten digit recognition systems under axial and lateral misalignments are reported in Figs. 2.2B and Fig. 2.2E, respectively. Figure 2.3 also provides a direct comparison of the blind inference accuracies of these two all-

optical diffractive machine learning architectures under different levels of misalignments. Figs. 2.3A and 2.3G compare the error-free designs of differential and standard diffractive network architectures, which reveal that although the differential design achieves slightly better blind inference accuracy, 97.93%, in the absence of alignment errors, as soon as the misalignments reach beyond a certain level, the performance of a differential design decreases faster than the standard diffractive network. This means that they are more vulnerable against the system variations that they were not trained against. Since the number of detectors inside an output region-of-interest is twice as many in differential diffractive networks compared to the standard diffractive network architecture (see Fig. 2.1A-B), the detector signals are more prone to have cross-talk when the diffractive layers are experiencing uncontrolled mechanical displacements. With the introduction of vaccination during the training phase, however, differential diffractive network models can adapt to these system variations as in the case of standard diffractive optical networks. Compared to standard diffractive optical networks, the differential counterparts that are vaccinated generate higher inference accuracies when the misalignment levels are small. In Fig. 2.3H, for instance, the vaccinated differential design (red curve) achieves 97.3% blind inference accuracy while the vaccinated standard diffractive network (blue curve) can provide 96.91% for the case $\Delta_{test} = \Delta_{tr} = 0.53\lambda$. In Fig. 2.3I, where the vaccination range on x and y axis is twice as large compared to Fig. 2.3H, the differential network reveals the correct digit classes with an accuracy of 96.18% when it is tested at an equal displacement/misalignment uncertainty to its vaccination level; on the other hand, the standard diffractive network can achieve 95.79% under the same training and testing conditions. Beyond this level of misalignment, the differential systems slowly lose their performance advantage and the standard diffractive networks starts to perform on par with their differential counterparts. One exception to this

behavior is shown in Fig. 2.3K, where the misalignment range of the diffractive layers during the training causes cross-talk among the differential detectors at a level that hurts the evolution of the differential diffractive network, leading to a consistently worse inference performance compared to the standard diffractive design. A similar effect also exists for the case illustrated in Fig. 2.3L; however, this time, the standard diffractive optical network design also experiences a similar level of cross-talk among the class detectors at the output plane. Therefore, as demonstrated in Fig. 2.3L, the differential diffractive optical network recovers its performance gain thriving over the standard diffractive network design with a higher optical classification accuracy. This performance gain of the differential design depicted in Fig. 2.3L, can be translated to the smaller misalignment cases, e.g., $\Delta_{test} = \Delta_{tr} = 4.24\lambda$, simply by increasing the distance between the detectors at the output plane for differential diffractive optical network designs, i.e. setting the region-of-interest covering the detectors to be larger compared to the standard diffractive network architecture.

Figure 2.3 also outlines a comparison of the differential and standard diffractive all-optical object recognition systems against hybrid diffractive neural networks under various levels of misalignments. For the hybrid neural network models presented here, we jointly trained a 5-layer diffractive optical front-end and a single-layer fully-connected electronic network, communicating through discrete detectors at the output plane. To provide a fair comparison with the all-optical diffractive systems, we used 10 discrete detectors at the output plane of these hybrid configurations, same as in the standard diffractive optical network designs (see Figs. 2.1A and 2.1C). The blind inference accuracies obtained by these hybrid neural network systems under different levels of misalignments are shown in Figs. 2.2C and 2.2F. When the opto-mechanical assembly of the diffractive network is perfect, the error-free, jointly-optimized

hybrid neural network architecture can achieve 98.3% classification accuracy surpassing the all-optical counterparts as well as the all-electronic performance of a single-layer fully-connected network, which achieves 92.48% classification accuracy using >75-fold more trainable connections without the diffractive optical network front-end. As the level of misalignments increases, however, the error-free hybrid network fails to maintain its performance and its inference accuracy quickly falls. The v-D^2NN framework helps the hybrid neural systems during the joint evolution of the diffractive and the electronic networks and makes them resilient to misalignments. For example, the handwritten digit classification accuracy values presented for the standard diffractive networks in Fig. 2.3H (96.91%) and Fig. 2.3I (95.79%) have improved to 97.92% and 97.15%, respectively, for the hybrid neural network system (yellow curve), indicating ~1% accuracy gain over the all-optical models under the same level of misalignment (i.e., 0.53λ for Fig. 2.3H and 1.06λ for Fig. 2.3I). As the level of misalignments in the diffractive optical front-end increases, the cross-talk between the detectors at the output plane also increases. However, for a hybrid network design there is no *direct* correspondence between the data classes and the output detectors, and therefore the joint-training under the vaccination scheme introduced in this work directs the evolution of the electronic network model accordingly and opens up the performance gap further between the all-optical diffractive classification networks and the hybrid systems as illustrated in Figs. 2.3K and 2.3L. A similar comparative analysis, along the lines of Figs. 2.2 and 2.3, is also conducted for phase-encoded input objects (Fashion-MNIST dataset), which is reported in Figs. 2.4 and 2.5.

Experimental results

The error-free standard diffractive network design that achieves 97.77% blind inference accuracy for the MNIST dataset as presented in Figs. 2.2A, 2.2D, 2.3A and 2.3G, offers a power efficiency of ~0.07% on average over the blind testing samples. This relatively low power efficiency is mostly due to the absorption of our 3D printing material at THz band. Specifically, ~88.62% of the optical power right after the object is absorbed by the 5 diffractive layers, while 11.17% is scattered around during the light propagation. Due to the limited optical power in our THz source and the noise floor of our detector, we trained an error-free standard diffractive optical network model with a slightly compromised digit classification performance for the experimental verification of our v-D^2NN framework. This new error-free diffractive network provides a blind inference accuracy of 97.19%, and it obtains ~3× higher power efficiency of ~0.2%. In addition to improved power efficiency, this new diffractive network model with 97.19% classification accuracy also achieves ~10× better signal contrast (ψ)[78] between the optical signal collected by the detector corresponding to the true object label and its closest competitor, i.e. the second maximum signal. The layers of this error-free diffractive network are shown in Fig. 2.1E. In addition, the comparison between the error-free, high-contrast standard diffractive optical network model and its lower contrast, lower efficiency counterpart in terms of their inference performance under misalignments is reported in Fig. 2.6A.

Following the same power-efficient design strategy, we trained another diffractive optical network that is *vaccinated* against *both the lateral and axial misalignments* with the training parameters (Δ_{tr}, $\Delta_{(z,tr)}$) taken as (4.24λ, 4.8λ). As in the case of the error-free design, the inference accuracy of this new vaccinated diffractive network shown in Fig. 2.7A is also compromised compared to the standard diffractive networks presented in Fig. 2.2D and Fig. 2.3K since it was trained to improve power efficiency and signal contrast. This design can

achieve 89.5% blind classification accuracy for handwritten digits under ideal conditions, with the diffractive layers reported in Fig. 2.1D. A comprehensive comparison of the blind inference accuracies of the vaccinated diffractive networks shown in Figs. 2.2 and 2.3 and their high-contrast, high-efficiency counterparts are reported in Fig. 2.6B.

The experimental verification of our v-D^2NN framework was based on the comparison of the vaccinated and the error-free standard diffractive optical network designs in terms of the accuracy of their optical classification decisions under inter-layer misalignments. To this end, we fabricated the diffractive layers of the non-vaccinated and the vaccinated networks shown in Figs. 2.1D-E using 3D printing. The fabricated diffractive networks are depicted in Figs. 2.7C-D. In addition, we fabricated 6 MNIST digits selected from the blind testing dataset that are numerically correctly classified by both the vaccinated and the non-vaccinated diffractive network models without any misalignments. For a fair comparison, we grouped the correctly classified handwritten digits based on the signal contrast statistics provided by the non-vaccinated design. With μ_{sc}, σ_{sc} denoting the mean and the standard deviation of the signal contrast generated by the error-free diffractive network over the correctly classified blind testing MNIST digits, we selected 2 handwritten digits (Set 1) that satisfies the condition $\mu_{sc}+\sigma_{sc} < \{\psi, \psi'\} < \mu_{sc}+2\sigma_{sc}$, where ψ and ψ' denote the signal contrasts created by the error-free and the vaccinated designs for a given input object, respectively. The condition on ψ and ψ' for the second set of 3D printed handwritten digits (Set 2), on the other hand, is slightly less restrictive, $\mu_{sc} < \{\psi, \psi'\} < \mu_{sc}+\sigma_{sc}$. By using this outlined approach, we selected 6 experimental test objects in total that are equally favorable for both the vaccinated and non-vaccinated diffractive networks.

To test the performance of the error-free and vaccinated diffractive network designs under different levels of misalignments, we shifted the 3rd layer of both diffractive systems to 12 different locations around its ideal location as depicted in Fig. 2.7B. The perturbed locations of the 3rd diffractive layer covers 4 different spots on each orthogonal direction. The distances between these locations are 1.2mm (1.6λ) along x and y, and 2.4mm (3.2λ) along z axes. These shifts cover a total length of 6.4λ (12 times the smallest feature size) along (x,y) and 12.8λ (0.32×40λ) along z axis, respectively.

Figure 2.7E shows a schematic of our THz setup that was used to test these diffractive networks and their misalignment performances. Figure 2.8 reports the experimentally obtained optical signals for a handwritten digit '0' from Set 1 and a handwritten digit '5' from Set 2, received by the class detectors at the output plane based on the 13 different locations of the 3rd diffractive layer of the vaccinated and the error-free networks. The first thing to note is that both the vaccinated and non-vaccinated networks can classify the two digits correctly when the 3rd layer is placed at its ideal location within the set-up. As illustrated in Fig. 2.8A, as we perturb the location of the 3rd layer, the error-free diffractive network fails at 9 locations while the vaccinated network correctly infers the object label at all the 13 locations for the handwritten digit '0'. In addition, the vaccinated network maintains its perfect record of experimental inference for the digit '5' despite the inter-layer misalignments as depicted in Fig. 2.8B. The error-free design, on the other hand, fails at 2 different locations of its 3rd layer misalignment (see Fig. 2.8B). The experimental results for the remaining 4 digits are presented in Figs. 2.9 and 2.10, confirming the same conclusions. In our experiments, all the objects were correctly classified when the 3rd layer was placed at its ideal location. Out of the remaining 72 measurements (6 objects × 12 shifted/misaligned locations of the 3rd layer), the error-free design

failed to infer the correct object class in 23 cases, while the vaccinated network failed only 2 times, demonstrating its robustness against a wide range of misalignments as intended by the v-D^2NN framework.

2.3 Discussion

As an example of a severe case of lateral misalignments, we investigated a scenario where each diffractive layer can move within the range (-8.48λ,8.48λ) around the optical axis in x and y directions. As demonstrated in Fig. 2.2D and Fig. 2.3G, when the error-free design (dark blue) is exposed to such large lateral misalignments, it can only achieve 12.8% test accuracy, i.e., it barely surpasses random guessing of the object classes. A diffractive optical network that is vaccinated against the same level of uncontrolled layer movement can partially recover the inference performance providing 67.53% blind inference accuracy. As the best performer, the hybrid neural network system composed of a 5-layer diffractive optical network and a single-layer fully-connected network can take this accuracy value up to 79.6% under the same level of misalignments, within the range (-8.48λ,8.48λ) for both x and y direction of each layer. When we compare the total allowed displacement range of each layer within the diffractive network (i.e., 16.96λ in each direction) and the size of our diffractive layers (106.66λ), we can see that they are quite comparable. If we imagine a lens-based optical imaging system and an associated machine vision architecture, in the presence of such serious opto-mechanical misalignments, this system would also fail due to acute aberrations substantially decreasing the image quality and the resolution. Our main motivation to include this severe misalignment case in our analyses was to test the limits of the adaptability of our vaccinated systems.

Figures 2.11A-B further summarize the inference accuracies of the differential diffractive networks and hybrid neural network systems at discrete points sampled from the corresponding curves depicted in Figs. 2.3G-L. In Fig. 2.11A, the best inference accuracy is achieved by the error-free (non-vaccinated) differential diffractive network model under perfect alignment of its layers. However, its performance drops in the presence of an imperfect opto-mechanical assembly. The vaccinated, diffractive all-optical classification networks provide major advantages to cope with the undesired system variations achieving higher inference accuracies despite misalignments. The joint-training of hybrid systems that are composed of a diffractive optical front-end and a single-layer electronic network (back-end) can adapt to uncontrolled mechanical perturbations achieving higher inference accuracies compared to all-optical image classification systems. These results further highlight that, operating with only a few discrete opto-electronic detectors at the output plane, the D^2NN-based hybrid architectures offer unique opportunities for the design of low-latency, power-efficient and memory-friendly machine vision systems for various applications.

On top of the translational layer-to-layer alignment errors, the presented framework can also be extended to accommodate 3D rotational misalignments of diffractive layers. While undesired in-plane rotations of diffractive layers can be readily addressed based on the 2D coordinate transformations performed through unitary rotation matrices incorporated into the optical forward model detailed, handling possible out-of-plane rotations of diffractive optical network layers requires further modifications to the formulation of wave propagation between tilted planes [89,90]. Beyond misalignments or displacements of diffractive layers, the presented vaccination framework can also be harnessed to decrease the sensitivity of diffractive optical networks to various error sources, e.g. detection noise or fabrication defects. At its core, the

presented framework can be interpreted as a training regularization method that avoids overfitting of a machine learning hardware to the specific 3D physical structure, distances and operational conditions, which are often assumed to be deterministic, precise and ideal during the training phase. In this respect, beyond its application to practically improve diffractive optical neural networks, the core principles introduced in our work can be extended to train other machine learning platforms [76,91,92] to mitigate various physical error sources that can cause deviations between the designed inference models and their corresponding physical implementations.

In conclusion, we presented a design framework that introduces the use of probabilistic layer-to-layer misalignments during the training of diffractive neural networks to increase their robustness against physical misalignments. Although the experimental demonstrations of our vaccinated design framework were conducted using THz wavelengths and 3D printed diffractive layers, the presented principles and methods can readily be applicable to diffractive optical networks that operate at other parts of the electromagnetic spectrum, including e.g., visible wavelengths. In fact, as the wavelength of operation gets smaller, the impact and importance of the presented framework will be better highlighted. We believe the presented training strategy will find use in the design of diffractive optical network-based machine vision and sensing systems, spanning different applications.

2.3 Materials and Methods

THz setup

The schematic diagram of the experimental setup is given in Fig. 2.7E. The THz wave incident on the object was generated through a horn antenna compatible with the source WR2.2

modular amplifier/multiplier chain (AMC) from Virginia Diode Inc. (VDI). Electrically modulated with 1 kHz square wave, the AMC received an RF input signal that is a 16 dBm sinusoidal waveform at 11.111 GHz (f_{RF1}). This RF signal is multiplied 36 times to generate the continuous-wave (CW) radiation at 0.4 THz, corresponding to ~0.75 mm in wavelength. The exit aperture of the horn antenna was placed ~60 cm away from the object plane of the 3D-printed diffractive optical network. At the output plane of the diffractive optical network, we 3D-printed an output aperture that has 10 openings, each with a size of 4.8 mm × 4.8 mm, defining the class detectors at their relative locations. The diffracted THz light at the output plane was collected with a single-pixel Mixer/AMC from Virginia Diode Inc. (VDI). A 10 dBm sinusoidal signal at 11.083 GHz was sent to the detector as local oscillator for mixing, and the down-converted signal was at 1GHz. The 10 openings representing the class detectors was scanned by placing the single-pixel detector on an XY stage that was built by combining two linear motorized stages (Thorlabs NRT100). The scanning step size was set to be 1 mm within each aperture opening. The down-converted signal of single-pixel detector at each scan location was sent to low-noise amplifiers (Mini-Circuits ZRL-1150-LN+) to amplify the signal by 80 dBm and a 1 GHz (+/-10 MHz) bandpass filter (KL Electronics 3C40-1000/T10-O/O) to clean the noise coming from unwanted frequency bands. Following the amplification, the signal was passed through a tunable attenuator (HP 8495B) and a low-noise power detector (Mini-Circuits ZX47-60), then the output voltage was read by a lock-in amplifier (Stanford Research SR830). The modulation signal was used as the reference signal for the lock-in amplifier and accordingly, we conducted a calibration by tuning the attenuation and record the lock-in amplifier readings. The lock-in amplifier readings at each scan location were converted to linear scale according to the calibration. The

class scores shown in Figs. 2.8-2.10, were computed as the sum of the calibrated and converted lock-in amplifier output at every scan step within the corresponding class detector opening.

The diffractive optical networks were fabricated using a 3D printer (Objet30 Pro, Stratasys Ltd.). Each 3D-printed diffractive optical network consisted of an input object, 5 diffractive layers and an output aperture array corresponding to the desired locations of the class detectors (see Fig. 2.1A). While the active modulation area of our 3D printed diffractive layers was 8 cm × 8 cm (106.66λ × 106.66λ), they were printed as light modulating insets surrounded by a uniform slab of printed material with a thickness of 0.9 mm. The total size of each printed layer was selected large enough to accommodate the introduced shifts on the 3rd diffractive layer location (for misalignment testing), with a total size of 12.8 cm × 12.8 cm.

The output aperture array and the 3D-printed MNIST digits were coated with aluminum except the openings and object features. Each aperture at the output plane is a square covering an area of 4.8 mm × 4.8 mm, matching the assumed size during the training. The size of the printed MNIST digits was 4 cm × 4cm sampled at a rate of 0.4 mm in both x and y directions, matching the training forward model. A 3D-printed holder was used to align the 3D printed input object, 5 diffractive layers and the output aperture. Around the location of the 3rd layer, the holder had additional spatial features that allowed us to move this diffractive layer to 13 different locations including the ideal one (see Fig. 2.7).

Forward optical model

In a diffractive optical network, each unit diffractive feature of a layer represents a complex-valued transmittance learned during the training process that optimizes the *thickness,* h, of the features based on the complex-valued refractive index of the 3D-fabrication material, $\tau = n$

+ jκ. The characterization of the printing material in a THz-TDS setup revealed the values of n and κ as 1.7227 and 0.031, respectively, for a monochromatic THz light at 400 GHz. Our formulation represents the complex-valued transmittance function of a diffractive feature on layer, l, at coordinates (x_q, y_q, z_l) as;

$$t(x_q, y_q, z_l) = \exp\left(-\frac{2\pi\kappa h(x_q, y_q, z_l)}{\lambda}\right)\exp\left(j(n - n_{air})\frac{2\pi h(x_q, y_q, z_l)}{\lambda}\right) \quad (2.3)$$

where $h(x_q, y_q, z_l)$, n_{air} and z_l denote the thickness of a given feature, refractive index of air and the axial location of the layer, l, respectively. From the Rayleigh-Sommerfeld theory of diffraction, we can interpret every diffractive unit on layer, l, at (x_q, y_q, z_l), as the source of a secondary wave, $w^l_q(x, y, z)$,

$$w_q^l(x, y, z) = \frac{z_l}{r^2}\left(\frac{1}{2\pi r} + \frac{1}{j\lambda}\right)\exp(\frac{j2\pi r}{\lambda}) \quad (2.4)$$

where $r = ((x-x_q)^2 + (y-y_q)^2 + (z-z_l)^2)^{0.5}$. Therefore, the complex field coming out of the q^{th} feature of $(l+1)^{th}$ layer, $u^{l+1}_q(x,y,z)$ can be written as;

$$u_q^{l+1}(x, y, z) = t(x_q, y_q, z_{l+1})w_q^{l+1}(x, y, z)\left(\sum\nolimits_{k\in l} u_k^l(x_q, y_q, z_{l+1})\right) \quad (2.5)$$

We sampled our diffractive fields and surfaces at a sampling interval of 0.4 mm that is equal to 0.53λ. The smallest diffractive feature size was also equal to 0.4 mm. The learnable thickness of each feature, h, was defined over an auxiliary variable, h_a;

$$h = Q_4(\frac{\sin(h_a) + 1}{2}(h_m - h_b)) + h_b \quad (2.6)$$

where h_m and h_b denote the maximum modulation thickness and base thickness, respectively. Taking h_b as 0.5 mm and h_m as 1 mm, we limited the printed thickness values between 0.5 mm and 1.5 mm. The minimum thickness h_b was used to mainly ensure the mechanical stability of the 3D printed layers against cracks and bending. The operator q(.) represents the quantization operator. We quantized the thickness values to 16 discrete levels (0.0625 mm per step). For the initialization of the diffractive layers at the beginning of the training, the thickness of each feature was taken as a uniformly distributed random variable between 0.9 mm and 1.1 mm, including the base thickness.

The training of the vaccinated diffractive optical networks follows the same optical forward model outlined in the previous section, except that it additionally introduces statistical variations following the models of the error sources in a diffractive network. The components of the 3D displacement vector of the l^{th} diffractive layer, $\mathbf{D^l} = (D^l_x, D^l_y, D^l_z)$, were defined as uniformly distributed random variables defined by Eq. (2.1). The vaccination strategy uses different sets of displacement vectors at every iteration (batch) to introduce undesired misalignments of the diffractive layers during the training. With $\mathbf{D}(l,i) = (D(l,i)_x, D(l,i)_y, D(l,i)_z)$ denoting the random displacement that the l^{th} layer experiences at i^{th} iteration, Eq. (2.6) was adjusted according to the longitudinal shift of the successive layers, $D_z(l,i)$ and $D_z(l+1,i)$, i.e., the light propagation distances between the diffractive layers were varied at every iteration. To implement the continuous lateral displacement of diffractive layers, we used:

$$t^{(l,i)}(x + D_x^{(l.i)}, y + D_y^{(l.i)}) = \int\int T^{(l,i)}(u,v)\exp\left(j2\pi\left(u\left(x + D_x^{(l.i)}\right) + v\left(y + D_y^{(l.i)}\right)\right)\right)dudv \quad (2.7)$$

where $t^{(l,i)}(x,y)$ denotes the 2-dimensional complex modulation function of layer l, at i^{th} iteration, and $T^{(l,i)}(u,v)$ represents its spatial Fourier transform defined over the 2D spatial frequency space (u,v).

Loss functions and class scores

In our forward training model, without loss of generality, we modeled our detectors as ratiometric sensors that capture the ratio of the optical power incident over their active area, $\mathbf{P_d}$, and the optical power incident over the object at the input plane, P_{obj}. Based on this, the optical signal **vector** collected by output detectors, $\mathbf{I_d}$, was formulated as:

$$\boldsymbol{I_d} = \frac{\boldsymbol{P_d}}{P_{obj}} \quad (2.8)$$

For all three diffractive object classification systems depicted in Fig. 2.1, the cost function was defined as the widely-known softmax-cross-entropy (SCE);

$$\boldsymbol{\mathcal{L}} = -\sum_{c=1}^{\boldsymbol{C}} g_c \log\left(\frac{exp(s_c)}{\sum_{c=1}^{C} exp(s_c)}\right) \quad (2.9)$$

where g_c, s_c and C denote the binary entry in the label vector, the computed class score for the data class, c, and the number of data classes in a given dataset (e.g., C=10), respectively.

For the standard diffractive optical network architecture shown in Fig. 2.1A, the number of class detectors, N_d, is equal to the number of data classes, C. In this scheme, the class score vector, **s**, was computed by:

$$\boldsymbol{s} = T\frac{\boldsymbol{I_d}}{max(\boldsymbol{I_d}) + \varepsilon} \tag{2.10}$$

where T and ε are constants, i.e., non-trainable hyperparameters used during the training phase. The multiplicative factor T was empirically set to be equal to 10 to generate artificial signal contrast at the input of softmax function for more efficient convergence of training. The constant ε, on the other hand, was used to regularize the power efficiency of the standard diffractive object recognition systems. In particular, the standard diffractive neural network models presented in Figs. 2.2A, 2.2D and 2.3, as well as in the Figs. 2.4A, 2.4D and 2.5, were trained by taking $\varepsilon = 10^{-4}$ which results in low power efficiency, η, and low signal contrast, ψ. The 3D-printed diffractive optical networks, on the other hand, were trained by setting $\varepsilon = 10^{-3}$ to circumvent the effects of the limited signal-to-noise ratio in our experimental system. Trained with a higher ε value, these diffractive networks offer slightly compromised blind testing accuracies while providing significantly improved power efficiency, η, and signal contrast, ψ, which are defined as:

$$\eta = I_{gt}\,,$$
$$\psi = I_{gt} - I_{sc}, \tag{2.11}$$

where I_{gt} and I_{sc} denote the optical signals measured by the class detector representing the ground truth label of the input object and its strongest competitor, i.e. the second maximum for a correctly classified input object, respectively. A comparison between the inference performances of low- and high- contrast variants of vaccinated and non-vaccinated standard diffractive optical networks under various levels of misalignments is presented in Fig. 2.6. As depicted in Fig. 2.6A, the high contrast, high efficiency standard diffractive networks are more robust against the undesired system variations/misalignments compared to their low-efficiency counterparts when both networks were trained under error-free conditions. Figure 2.6B, on the other hand, compares the standard diffractive network architectures that were tested within the same misalignment range used in their training. In this case, the low-contrast, power inefficient diffractive networks show their higher inference capacity advantage and adapt to the misalignments more effectively than the diffractive classification systems trained to favor higher power efficiency.

In a differential diffractive optical network system, the number of detectors is doubled, i.e. $N_d=2C$, where each pair represents the negative, $\mathbf{I_{d-}}$, and positive signal vector, $\mathbf{I_{d+}}$, contributing to the normalized differential signal, $\mathbf{I_{(d,n)}}$ (see Fig. 2.1B) defined as:

$$\boldsymbol{I}_{(\boldsymbol{d},\boldsymbol{n})} = \frac{\boldsymbol{I}_{\boldsymbol{d}+} - \boldsymbol{I}_{\boldsymbol{d}-}}{\boldsymbol{I}_{\boldsymbol{d}+} + \boldsymbol{I}_{\boldsymbol{d}-}} \tag{2.12}$$

In parallel, the class scores of a differential diffractive object classification system, **s**, are calculated by replacing the optical signal vector, $\mathbf{I_d}$, in Eq. (2.10) with the normalized differential signals, $\mathbf{I_{(d,n)}}$, depicted in Eq. (2.12).

Once the training is completed, these equations are not used in the numerical and experimental blind testing, meaning that *the class decision is made solely based on max($\mathbf{P_d}$) and max($\mathbf{P_{(d,n)}}$) in standard and differential diffractive network systems, respectively*.

In the hybrid neural network models, we jointly-trained 5-layer diffractive optical networks with an electronic network that has a single-layer fully-connected network with only 110 (100 multiplicative weights + 10 bias) trainable parameters. During the joint-evolution of these two neural networks, we normalized the optical signal collected by the detectors, $\mathbf{I_d}$, as depicted in Eq. (2.10) with T = 1. These normalized detector signals were then fed into the subsequent fully-connected layer in the electronic domain to compute the class scores, **s**, which was used in Eq. (2.9) for computing the classification loss before the error-backpropagation through both the electronic and diffractive optical networks.

Other details of training

All network models used in this work were trained using Python (v3.6.5) and TensorFlow (v1.15.0, Google Inc.). We selected Adam optimizer during the training of all the models, and its parameters were taken as the default values in TensorFlow and kept identical in each model. The learning rates of the diffractive optical networks and the electronic neural network were set to be 0.001 and 0.0002, respectively. The data of handwritten digits and fashion-products were both divided into three parts: training, validation and testing, containing 55K, 5K and 10K images, respectively. All object recognition systems were trained for 50 epochs with a batch size of 50 and the best model was selected based on the highest classification performance on the validation dataset. In the training of MNIST digits, the image information was encoded in the amplitude

channel at the object plane, while the Fashion-MNIST objects was assumed to be phase-only targets with their gray levels mapped to phase values between 0 and π.

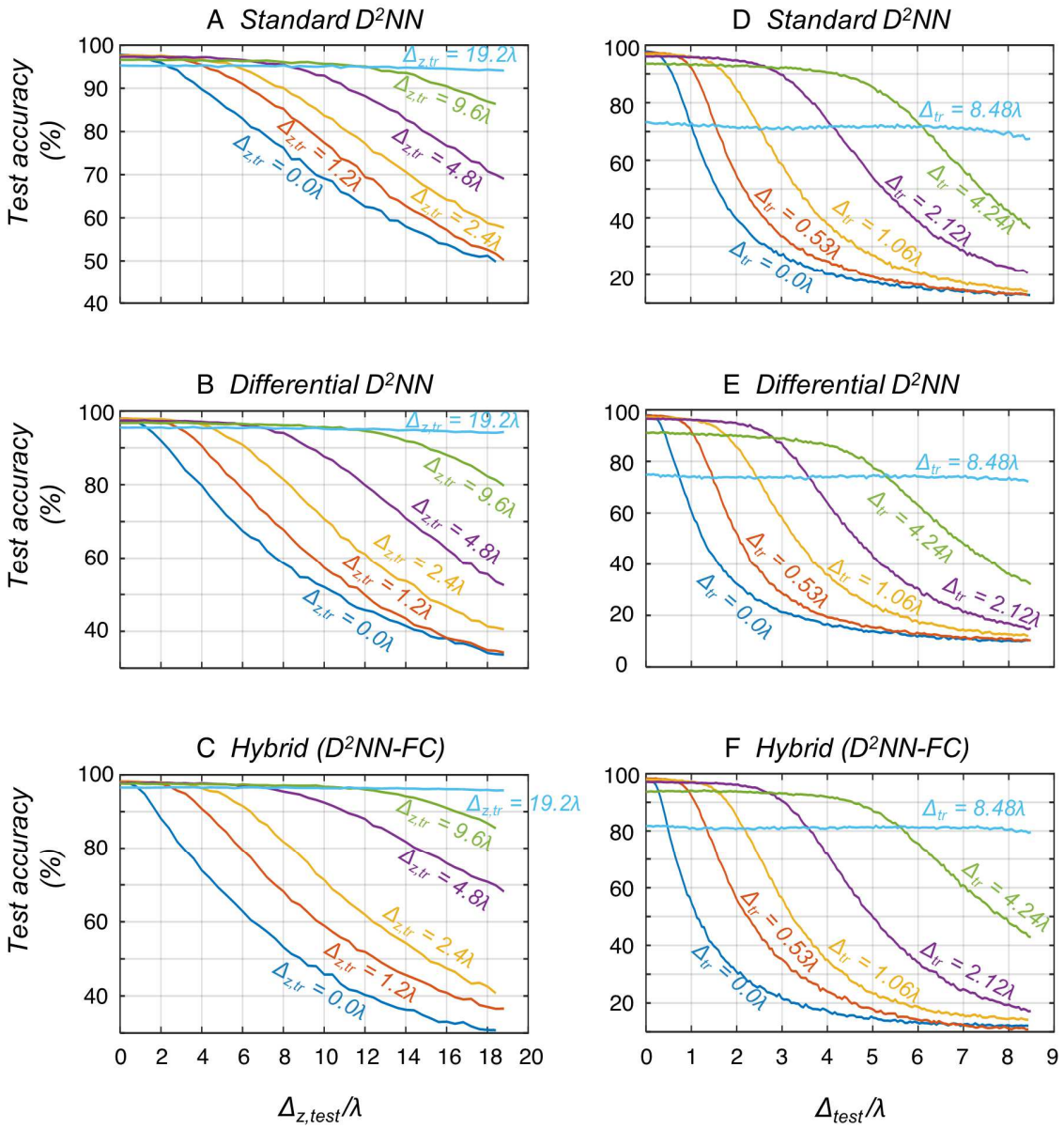

Fig. 2.2 The sensitivity of the blind inference accuracies of different types of D2NN-based object classification systems against various levels of misalignments. A Standard D^2NN systems trained for all-optical handwritten digit classification with and without vaccination were tested against various levels of *axial* misalignments, determined by $\Delta_{Z,test}$. B Same as A, except for differential D^2NN architectures. C Same as A and B, except for hybrid (D^2NN-FC) systems comprised of a jointly-trained 5-layer D^2NN optical front-end and a single-layer fully-connected neural network at the electronic back-end, combined through 10 discrete opto-electronic detectors (see Fig. 2.1C). The comparison of these blind testing results reveals that as the axial misalignment increases during the training, $\Delta_{Z,tr}$, the inference accuracy of these machine vision systems decrease slightly but at the same time they are able to maintain their performance over a wider range of misalignments during the blind testing, $\Delta_{Z,test}$. D Standard D^2NN systems trained for all-optical handwritten digit recognition with and without vaccination were tested against various levels of *lateral* misalignment levels, determined by Δ_{test}. E Same as D except for differential D^2NNs architectures. F Same as E and F, except for hybrid object recognition systems comprised of a jointly-trained 5-layer D^2NN optical front-end and a single-layer fully-connected neural network at the electronic back-end, combined through 10 discrete opto-electronic detectors. The proposed vaccination-based training strategy improves the resilience of these diffractive networks to uncontrolled *lateral* and *axial* displacements of the diffractive layers with a modest compromise of the inference performance depending on the misalignment range used in the training phase.

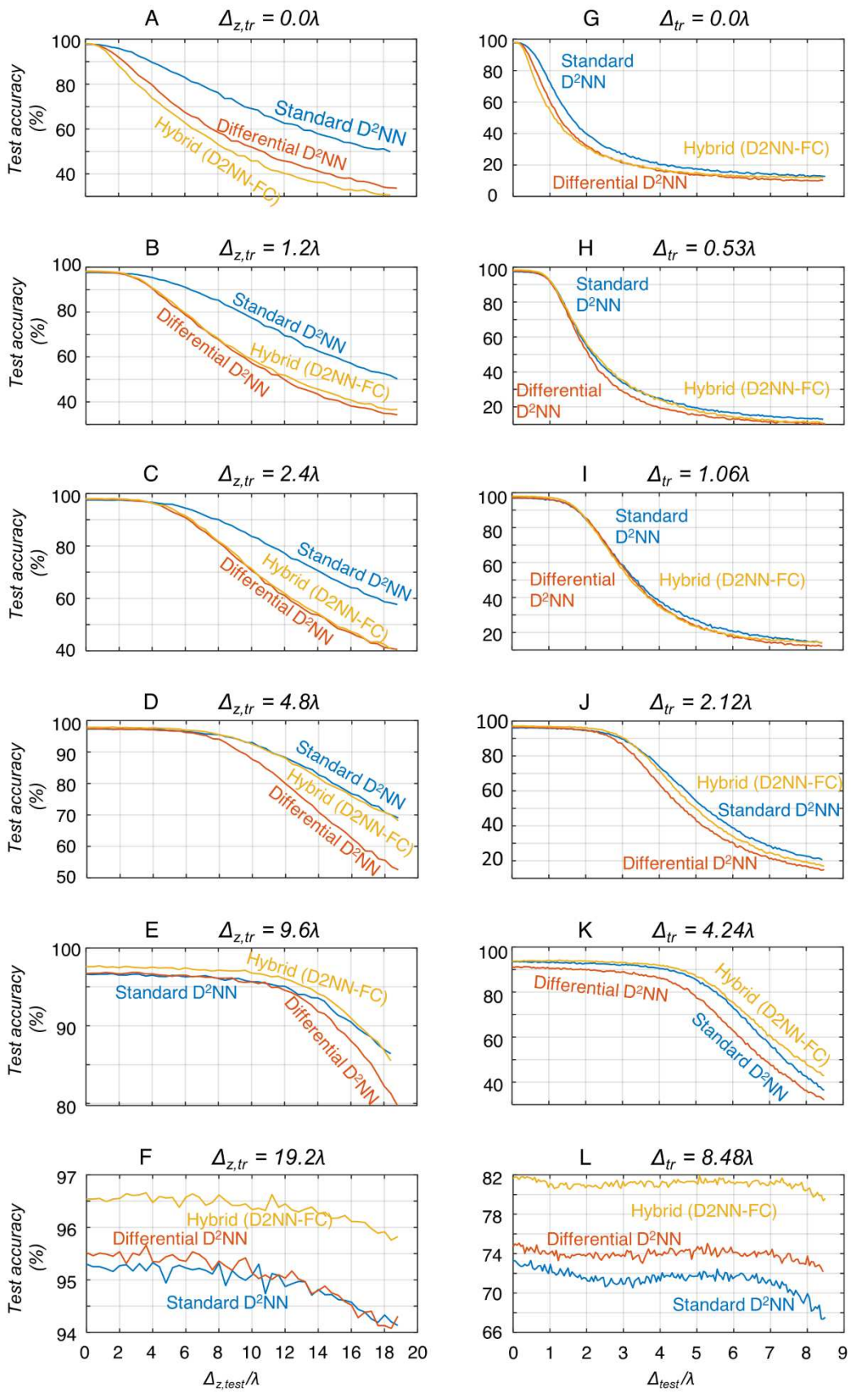
A
$\Delta_{z,tr}$ = 0.0λ
Test accuracy (%)
Standard D²NN
Differential D²NN
Hybrid (D2NN-FC)
B
$\Delta_{z,tr}$ = 1.2λ
C
$\Delta_{z,tr}$ = 2.4λ
D
$\Delta_{z,tr}$ = 4.8λ
E
$\Delta_{z,tr}$ = 9.6λ
F
$\Delta_{z,tr}$ = 19.2λ
$\Delta_{z,test}/\lambda$
G
Δ_{tr} = 0.0λ
H
Δ_{tr} = 0.53λ
I
Δ_{tr} = 1.06λ
J
Δ_{tr} = 2.12λ
K
Δ_{tr} = 4.24λ
L
Δ_{tr} = 8.48λ
Δ_{test}/λ

Fig. 2.3 Comparison of different types of D2NN-based object classification systems trained with the same range of misalignments. A Comparison of error-free designs, $\Delta_{Z,tr} = 0.0\lambda$, for standard (blue), differential (red) and hybrid (yellow) object classification systems against different levels of *axial* misalignments, $\Delta_{Z,test}$. B Comparison of standard (blue), differential (red) and hybrid (yellow) object classification systems against different levels of *axial* misalignments when they are trained with $\Delta_{Z,tr} = 1.2\lambda$. C,D,E and F are same as B, except during the training of the diffractive models the axial misalignment ranges are determined by $\Delta_{Z,tr}$, taken as 2.4λ, 4.8λ, 9.6λ and 19.2λ, respectively. G Comparison of error-free designs, $\Delta_{tr} = 0.0\lambda$, for standard (blue), differential (red) and hybrid (yellow) object recognition systems against different levels of *lateral* misalignments, Δ_{test}. H Comparison of standard (blue), differential (red) and hybrid (yellow) object classification systems against different levels of *lateral* misalignments when they are trained with $\Delta_{tr} = 0.53\lambda$. I,J,K and L are same as H, except the *lateral* misalignment ranges during the training are determined by Δ_{tr}, taken as 1.06λ, 2.12λ, 4.24λ and 8.48λ, respectively.

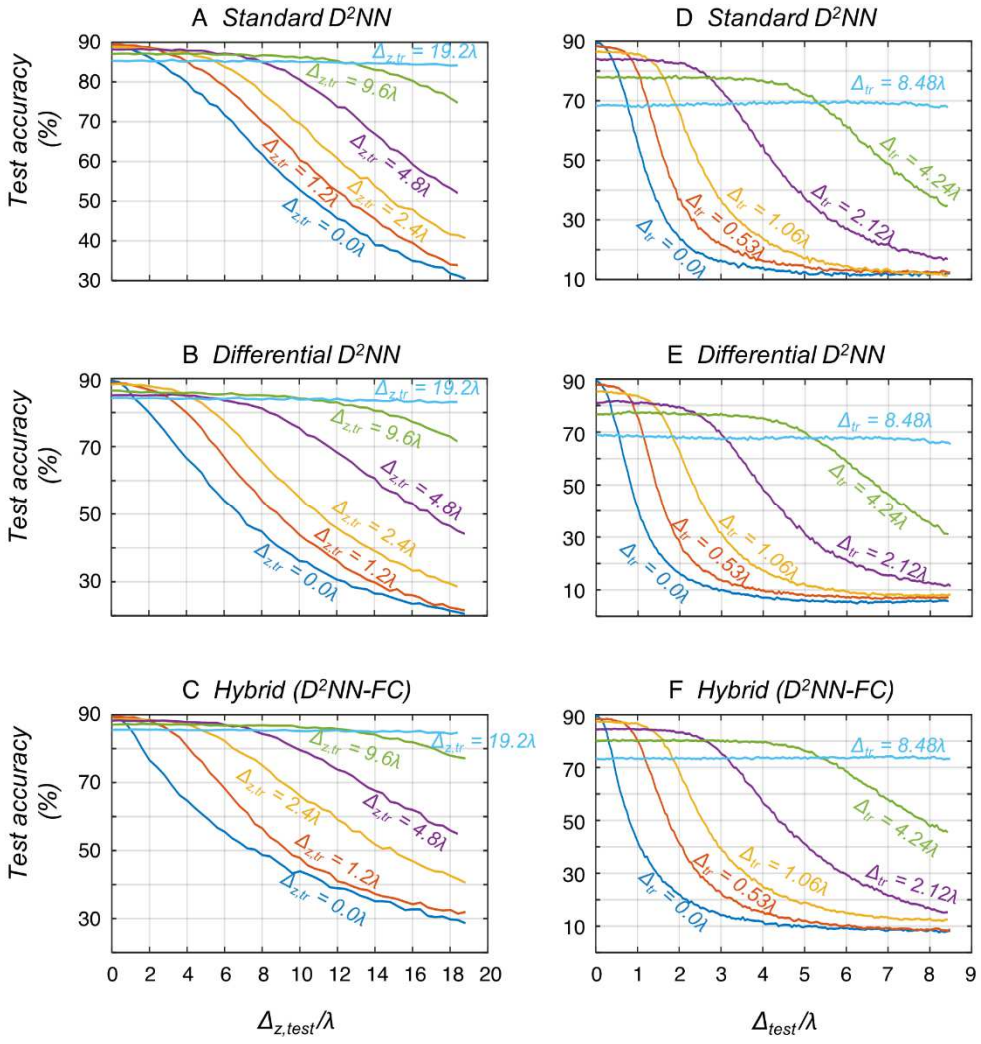

Fig. 2.4 The blind inference accuracies achieved by standard, differential and hybrid diffractive network systems for the classification of phase-encoded Fashion-MNIST images. Same as the Figure 2.2, except, the image dataset is Fashion-MNIST. Unlike amplitude encoded MNIST images at the input plane, the fashion products were assumed to represent phase-only targets at the object/input plane with their phase values restricted between 0 and π.

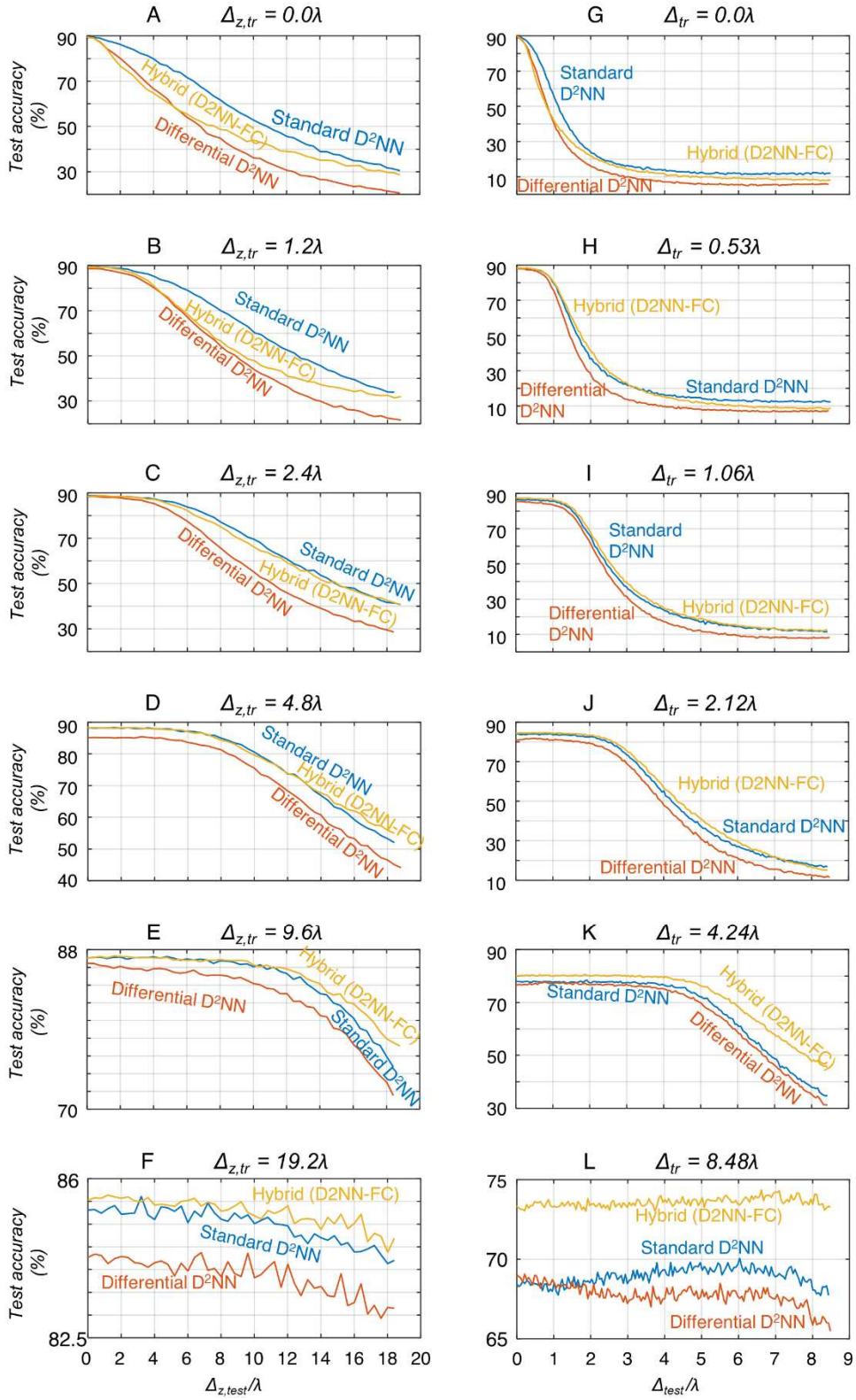
A
$\Delta_{z,tr}$ = 0.0λ
B
$\Delta_{z,tr}$ = 1.2λ
C
$\Delta_{z,tr}$ = 2.4λ
D
$\Delta_{z,tr}$ = 4.8λ
E
$\Delta_{z,tr}$ = 9.6λ
F
$\Delta_{z,tr}$ = 19.2λ
G
Δ_{tr} = 0.0λ
H
Δ_{tr} = 0.53λ
I
Δ_{tr} = 1.06λ
J
Δ_{tr} = 2.12λ
K
Δ_{tr} = 4.24λ
L
Δ_{tr} = 8.48λ
Test accuracy (%)
Standard D²NN
Hybrid (D2NN-FC)
Differential D²NN
$\Delta_{z,test}/\lambda$
Δ_{test}/λ

Fig. 2.5 Direct comparison of blind inference accuracies achieved by standard, differential and hybrid diffractive network systems for the classification of phase-encoded fashion products. Same as the Figure 2.3, except, the image dataset is Fashion-MNIST. Unlike amplitude encoded MNIST images at the input plane, the fashion products were assumed to represent phase-only targets at the object/input plane with their phase values restricted between 0 and π.

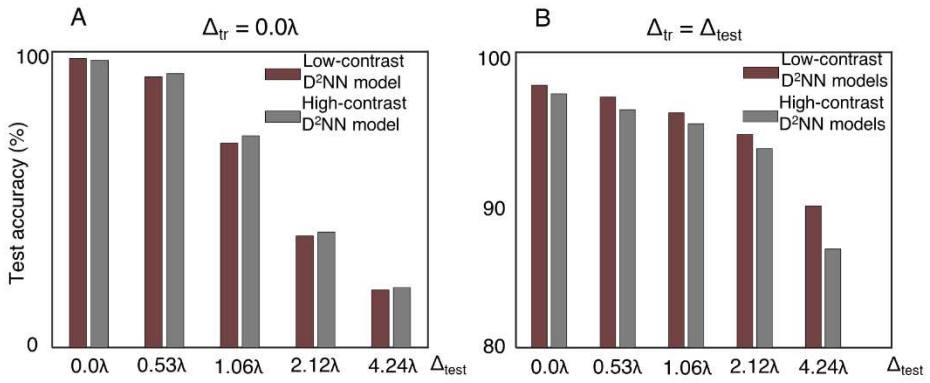

Fig. 2.6 The comparison between the low-contrast and high-contrast standard diffractive optical networks. A The inference accuracy values of two error-free standard optical network designs are compared. The low-contrast standard diffractive optical network (red) achieves slightly higher inference accuracy when the alignment is perfect. The high-contrast diffractive optical network, on the other hand, is slightly more robust against misalignments. B Trained with the v-D^2NN framework, low-contrast models use their higher inference capacity to adapt to misalignments, consistently achieving higher classification accuracies when they are tested under misalignment.

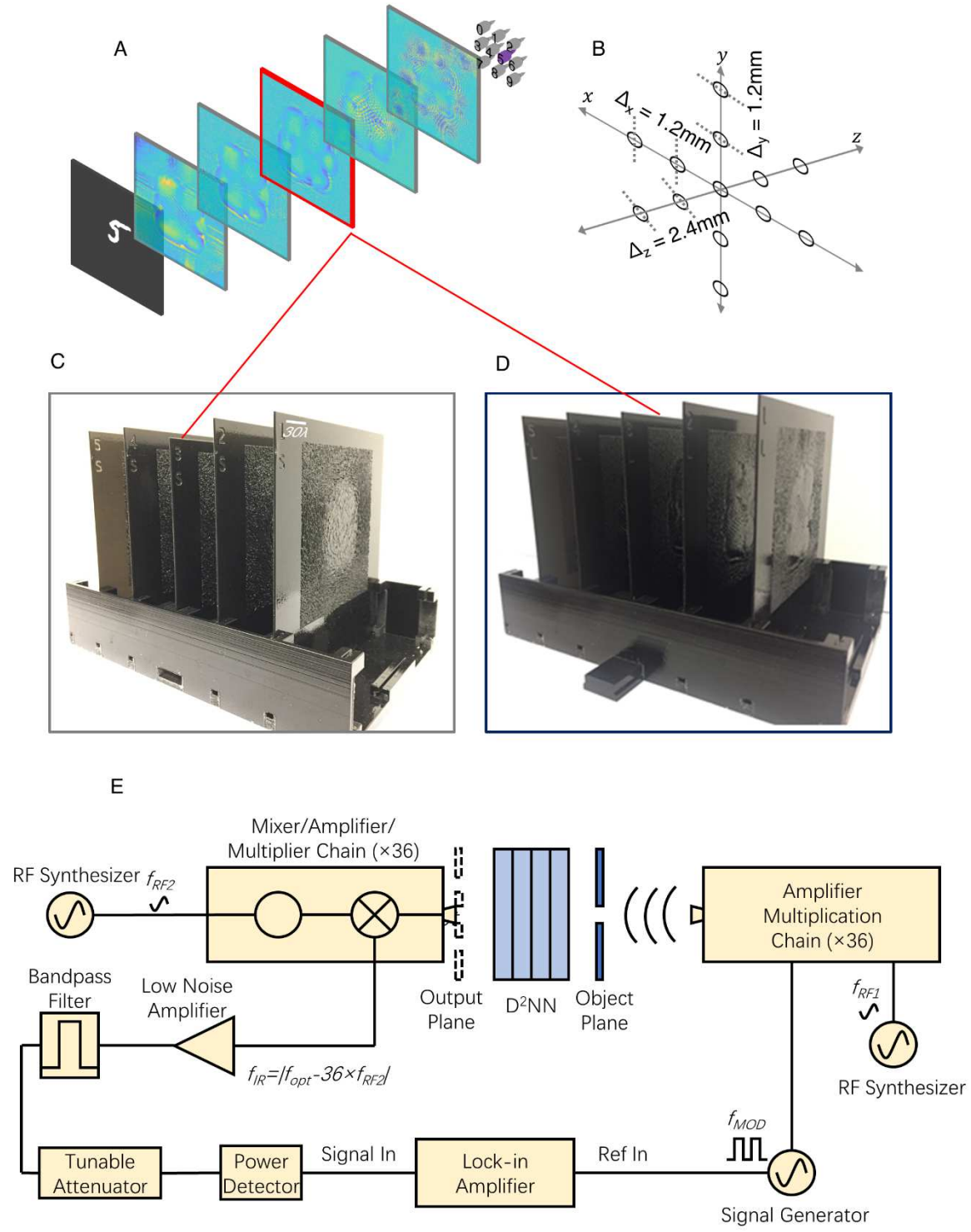

Fig. 2.7 Experimental testing of v-D2NN framework. A A diffractive optical network that is vaccinated against misalignments. This network is vaccinated against *both* lateral, $\Delta_{tr} = 4.24\lambda$, and axial, $\Delta_{z,tr} = 4.8\lambda$, misalignments. B

The location of the 3rd diffractive layer was on purpose altered throughout our measurements. Except the central location, the remaining 12 spots induce an inter-layer misalignment. C The 3D printed error-free design shown in Fig. 2.1E. D The 3D printed vaccinated design shown in A and Fig. 2.1D. E The schematic of the experimental setup.

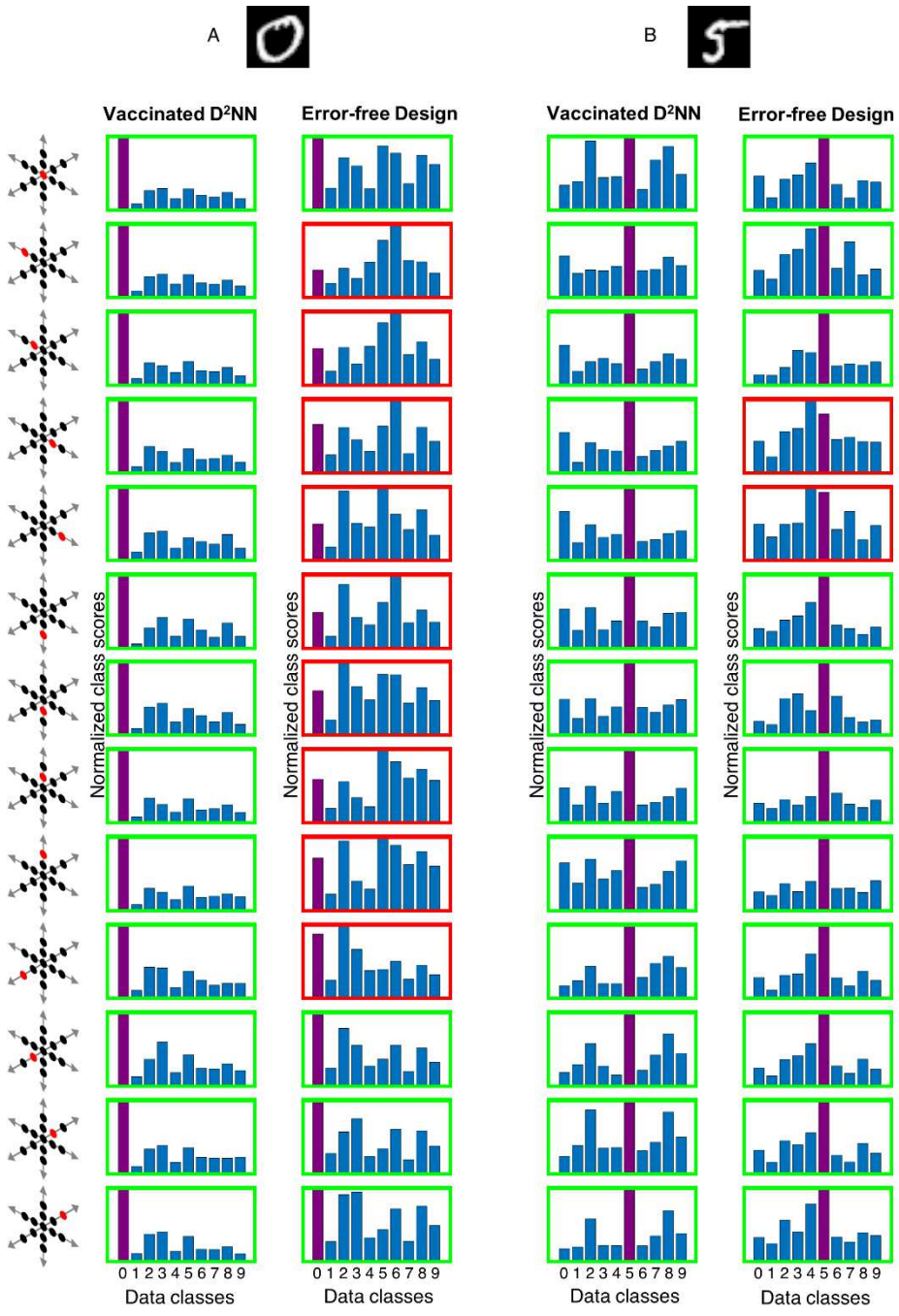

Fig. 2.8 Experimental image classification results as a function of misalignments. A The experimentally measured class scores for handwritten digit '0' selected from Set 1. B Same as A, except the input object is now a handwritten digit '5' selected from Set 2. The red dot within the coordinate system shown on the left-hand side represents the physical misalignment for each case (see Fig. 2.7B). Red (green) rectangles mean incorrect (correct) inference results.

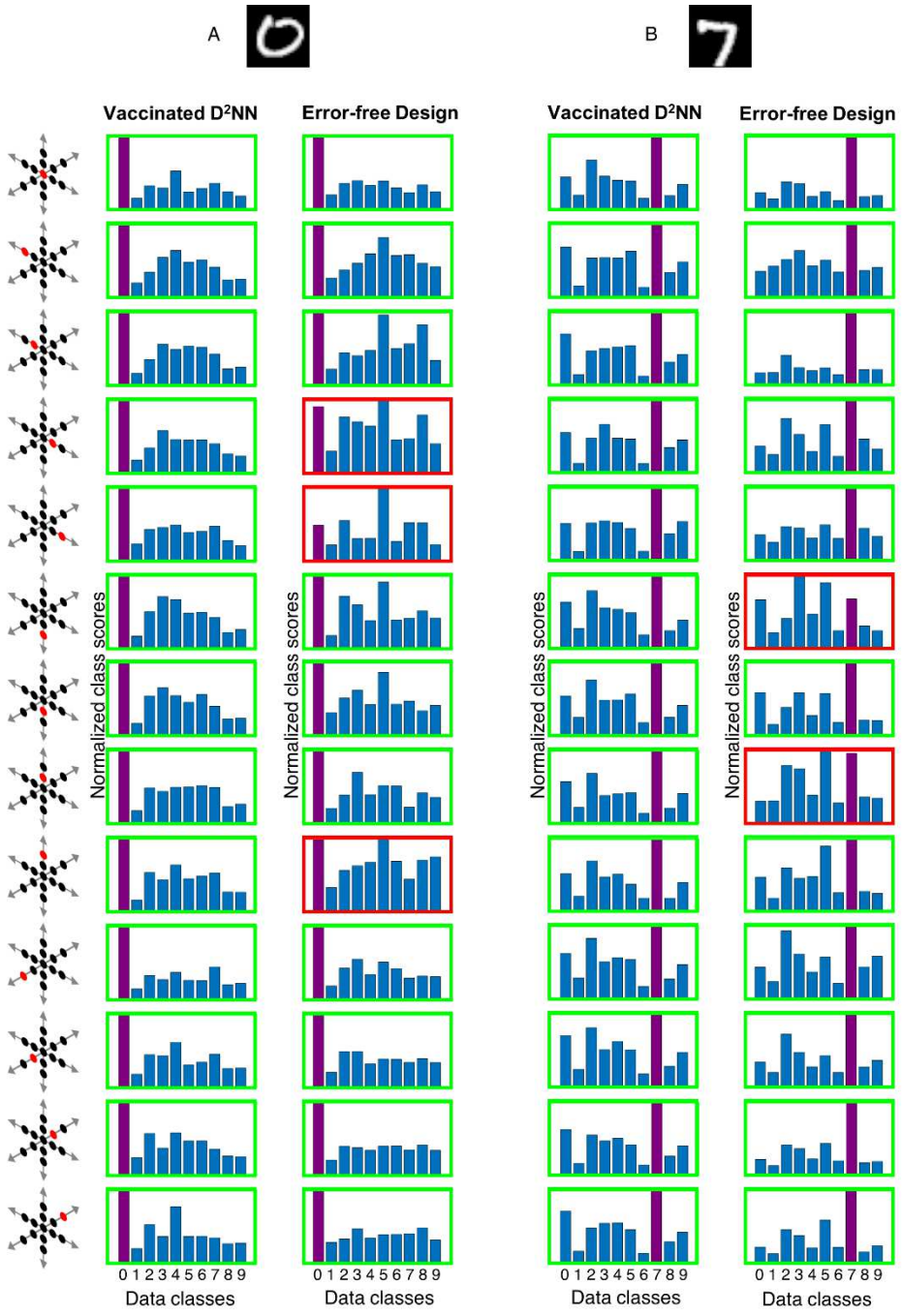

Fig. 2.9 Experimental image classification results as a function of misalignments. A The experimentally measured class scores for handwritten digit '0' selected from Set 2. B Same as A, except the input object is now a handwritten digit '7' selected from Set 2. The red dot within the coordinate system shown on the left-hand side represents the physical misalignment for each case. Red (green) rectangles mean incorrect (correct) inference results.

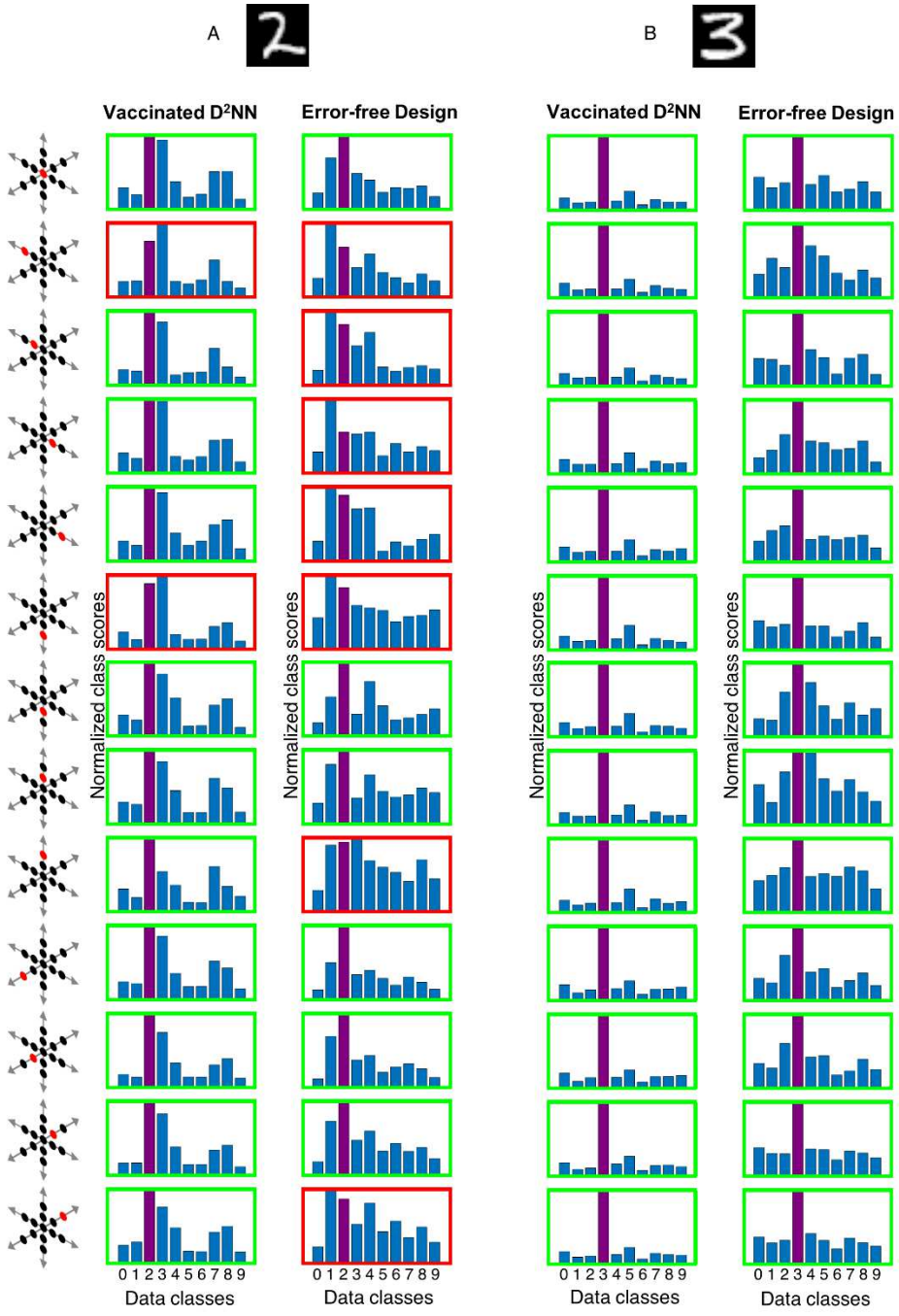

Fig. 2.10 Experimental image classification results as a function of misalignments. A The experimentally measured class scores for handwritten digit '2' selected from Set 1. B Same as A, except the input object is now a handwritten digit '3' selected from Set 2. The red dot within the coordinate system shown on the left-hand side represents the physical misalignment for each case. Red (green) rectangles mean incorrect (correct) inference results.

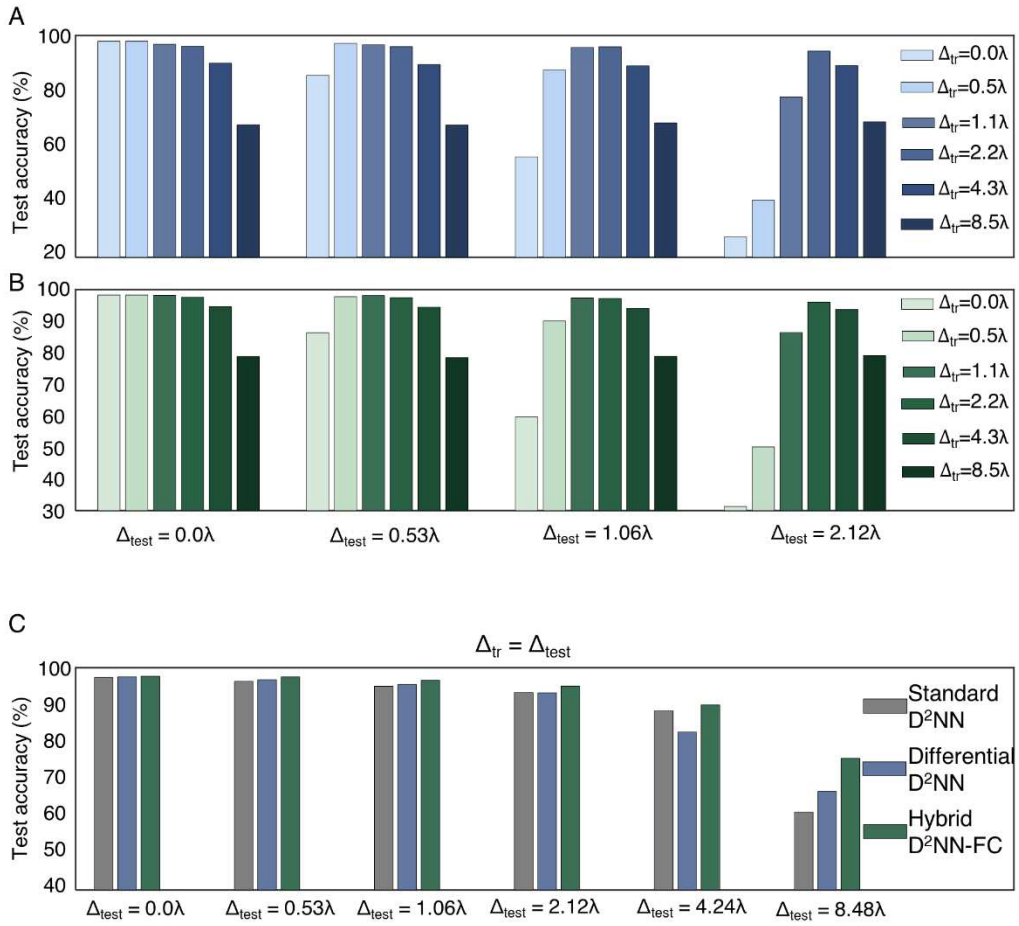

Fig. 2.11 Summary of the numerical results for vaccinated D2NNs. A The inference accuracy of the non-vaccinated ($\Delta_{tr} = 0.0\lambda$) and the vaccinated ($\Delta_{tr} > 0.0\lambda$) differential D^2NN systems trained for all-optical handwritten digit recognition quantified at different levels of testing misalignment ranges. The v-D^2NN framework allows the all-optical classification systems to preserve their inference performance over a large range of misalignments. B Same as A, except for hybrid (D^2NN-FC) systems comprised of a jointly-trained 5-layer D^2NN optical front-end and a single-layer fully-connected neural network at the electronic back-end combined through 10 discrete opto-electronic detectors (see Fig. 2.1C). C Vaccination comparison of 3 diffractive network-based machine learning architectures depicted in Fig. 2.1; $\Delta_{tr} = \Delta_{test}$.

Chapter 3 Scale-, Shift- and Rotation-Invariant Diffractive Optical Networks

Parts of this chapter have previously been published in D. Mengu et al. “Scale-, Shift- and Rotation-Invariant Diffractive Optical Networks” Scientific Reports, DOI: /10.1021/acsphotonics.0c01583. In this chapter, I will introduce diffractive optical classification networks that shows invariant inference accuracy under random scaling, translation and rotation of the input objects.

Recent research efforts in optical computing have gravitated towards developing optical neural networks that aim to benefit from the processing speed and parallelism of optics/photonics in machine learning applications. Among these endeavors, Diffractive Deep Neural Networks (D^2NNs) harness light-matter interaction over a series of trainable surfaces, designed using deep learning, to compute a desired statistical inference task as the light waves propagate from the input plane to the output field-of-view. Although, earlier studies have demonstrated the generalization capability of diffractive optical networks to unseen data, achieving e.g., >98% image classification accuracy for handwritten digits, these previous designs are in general sensitive to the spatial scaling, translation and rotation of the input objects. Here, we demonstrate a new training strategy for diffractive networks that introduces input object translation, rotation and/or scaling during the training phase as uniformly distributed random variables to build resilience in their blind inference performance against such object transformations. This training strategy successfully guides the evolution of the diffractive optical network design towards a solution that is scale-, shift- and rotation-invariant, which is especially important and useful for

dynamic machine vision applications in e.g., autonomous cars, in-vivo imaging of biomedical specimen, among others.

3.1 Introduction

Motivated by the success of deep learning[55,56] in various applications[44–48,50,52–54,58,66,68,93,94], optical neural networks have gained an important momentum in recent years. Although optical neural networks and related optical computing hardware are relatively at an earlier stage in terms of their inference and generalization capabilities, when compared to the state-of-the-art electronic deep neural networks and the underlying digital processors, optics/photonics technologies might potentially bring significant advantages for machine learning systems in terms of their power efficiency, parallelism and computational speed[20,22,60,69–71,74,75,81,84,85,92,95,96]. Among different physical architectures used for the design of optical neural networks[20,69–71,77,84,85,97], Diffractive Deep Neural Networks (D^2NNs)[77,79,98,78,99–101,80] utilize the diffraction of light through engineered surfaces/layers to form an optical network that is based on light-matter interaction and free-space propagation of light. D^2NNs offer a unique optical machine learning framework that formulates a given learning task as a black-box function approximation problem, parameterized through the trainable physical features of matter that control the phase and/or amplitude of light. One of the most convenient methods to devise a D^2NN is to employ multiple transmissive and/or reflective diffractive surfaces/layers that collectively form an optical network between an input and output field-of-view. During the training stage, the complex-valued transmission/reflection coefficients of the layers of a D^2NN are designed for a given statistical (or deterministic) task/goal, where each diffractive feature (i.e., neuron) of a given layer is iteratively adjusted during the training phase using e.g., the error back-propagation method[73,96,102]. After this training and design phase, the resulting diffractive layers/surfaces are physically fabricated using e.g., 3D printing or lithography, to form a passive optical network that performs inference as the input light diffracts from the

input plane to the output. Alternatively, the final diffractive layer models can also be implemented by using various types of spatial light modulators (SLMs) to bring reconfigurability and data adaptability to

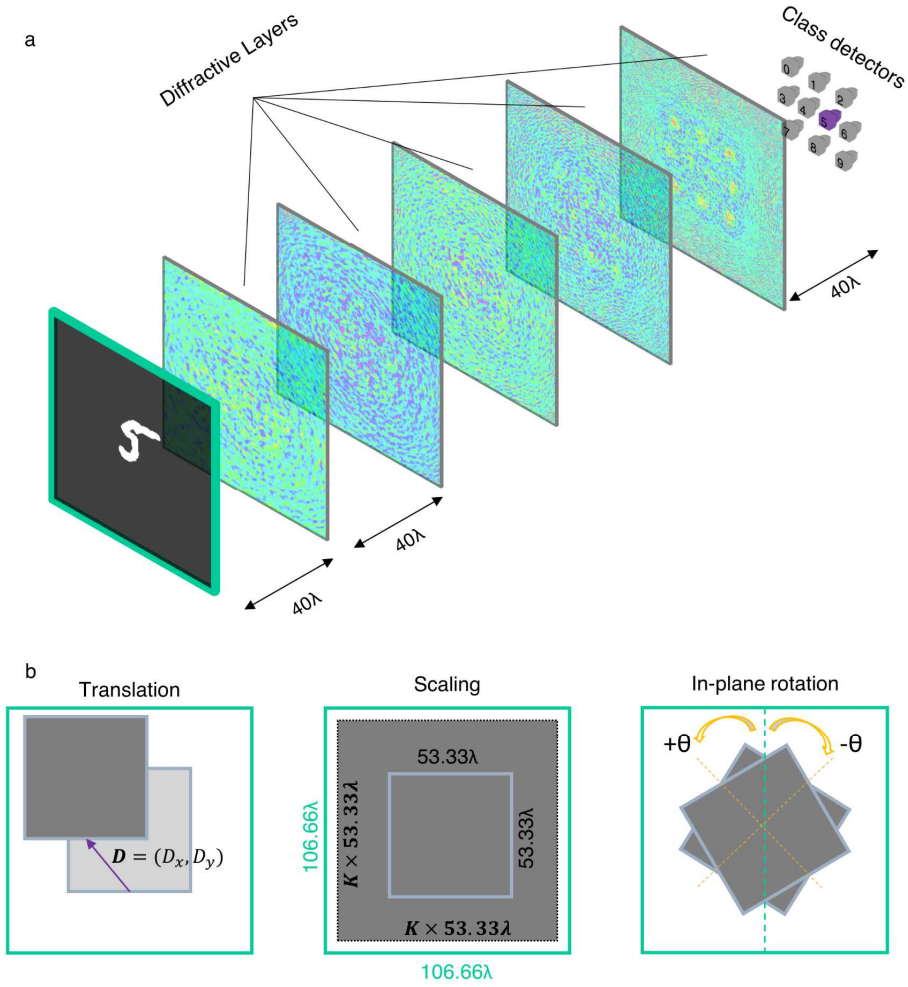

Fig. 3.1 Optical architecture of an all-optical diffractive classifier and geometric object transformations. (a) The layout of the diffractive optical networks trained and tested in this study. (b) The object transformations modeled during the training and testing of the diffractive optical networks presented in this chapter.

the diffractive network, at the expense of e.g., increased power consumption of the system.

Since the initial experimental demonstration of image classification using D^2NNs that are composed of 3D-printed diffractive layers[77,99], the optical inference capacity of diffractive optical networks has been significantly improved based on e.g., differential detection scheme, class-specific

designs and ensemble-learning techniques[79,98]. Owing to these systematic advances in diffractive optical networks and training methods, recent studies have reported classification accuracies of >98%, >90% and >62% for the datasets of handwritten digits (MNIST), fashion products (Fashion-MNIST) and CIFAR-10 images, respectively.[79,98] Beyond classification tasks, diffractive networks were also shown to serve as trainable optical front-ends, forming hybrid (optical-electronic) machine learning systems[78]. Replacing the conventional imaging-optics in machine vision systems with diffractive optical networks has been shown to offer unique opportunities to lower the computational complexity and burden on back-end electronic neural networks as well as to mitigate the inference accuracy loss due to pixel-pitch limited, low-resolution imaging systems.[78] Furthermore, in a recent study, diffractive optical networks have been trained to encode the spatial information of input objects into the power spectrum of the diffracted broadband light, enabling object classification and image reconstruction using only a single-pixel spectroscopic detector at the output plane, demonstrating an unconventional, task-specific and resource-efficient machine vision platform.[99] The extension of the diffractive optical networks and the related training forward models to conduct inference based on broadband light sources exhibits their potential in processing object information at multiple spectral bands simultaneously, e.g. red, green and blue channels of CIFAR-10 images in the visible.

In all of these existing diffractive optical network designs, the inference accuracies are in general sensitive to object transformations such as e.g., lateral translation, rotation, and/or scaling of the input objects that are frequently encountered in various machine vision applications. In this work, we quantify the sensitivity of diffractive optical networks to these uncertainties associated with the lateral position, scale and in-plane orientation/rotation angle of the input objects (see Fig. 3.1). Furthermore, we demonstrate a D^2NN design scheme that formulates these object transformations through random variables used during the deep learning-based training phase of the diffractive layers. In this manner, the

evolution of the layers of a diffractive optical network can adapt to random translation, scaling and rotation of the input objects and, hence, the blind inference capacity of the optical network can be maintained despite these input object uncertainties. The presented training strategy will enable diffractive optical networks to find applications in machine vision systems that require low-latency as well as memory- and power-efficient inference engines for monitoring dynamic events. Beyond diffractive networks, the outlined training scheme can be utilized in other optical machine learning platforms as well as in deep learning-based inverse design problems to create robust solutions that can sustain their target performance against undesired/uncontrolled input field transformations.

3.2 Results and Discussion

In a standard D^2NN-based optical image classifier[77–79,98,103], the number of opto-electronic detectors positioned at the output plane is equal to the number of classes in the target dataset and, each detector uniquely represents one data class (see Fig. 3.1a). The final class decision is based on the *max* operation over the collected optical signals by these class detectors. According to the diffractive network layout illustrated in Fig. 3.1a, the input objects (e.g., handwritten MNIST digits) lie within a pre-defined field-of-view (FOV) of $53.33\lambda \times 53.33\lambda$, where λ denotes the wavelength of the illumination light. The center of the FOV coincides with the optical axis passing through the center of the diffractive layers. The size of each diffractive layer is chosen to be $106.66\lambda \times 106.66\lambda$, i.e., exactly 2× the size of the input FOV on each lateral axis. The smallest diffractive feature size on each D^2NN layer is set to be $\sim 0.53\lambda$, i.e., there are 200×200 trainable features on each diffractive layer of a given D^2NN design. At the output plane, each detector is assumed to cover an area of $6.36\lambda \times 6.36\lambda$ and they are located within an output FOV of $53.33\lambda \times 53.33\lambda$ – matching the input FOV size.

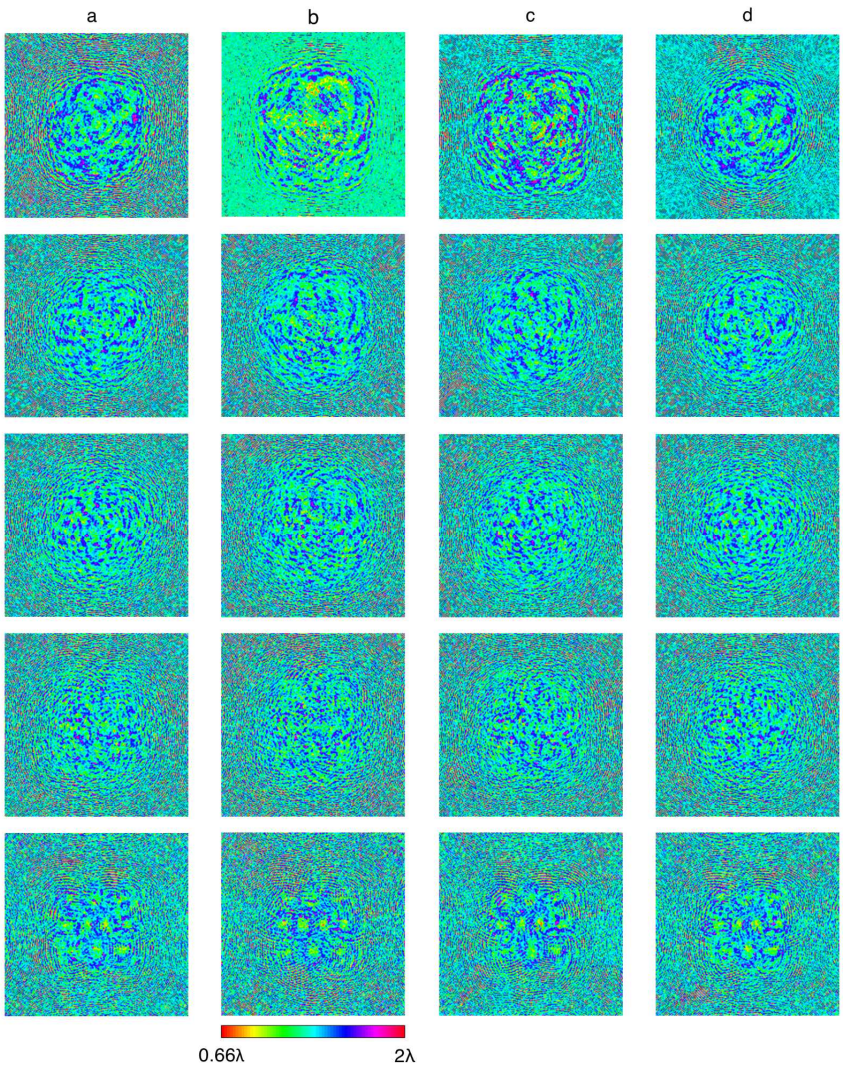

Fig. 3.2 The thickness profiles of the designed diffractive layers constituting (a) the standard design ($\Delta_{tr} = 0$); (b) the shift-invariant design trained with $\Delta_{tr} = 8.48\lambda$ (purple curve shown in Fig. 3.3); (c) the scale-invariant design trained with $\zeta_{tr} = 0.4$ (purple curve shown in Fig. 3.5); (d) the rotation-invariant design trained with $\theta_{tr} = 20°$ (purple curve shown in Fig. 3.6).

Based on these design parameters, a 5-layer diffractive optical network with phase-only modulation at each neuron achieves a blind testing accuracy of 97.64% for the classification of amplitude-encoded MNIST images illuminated with a uniform plane wave. Figure 3.2a

illustrates the thickness profiles of the resulting 5 diffractive layers, constituting this standard D^2NN design. To quantify the sensitivity of the blind inference accuracy of this D^2NN design against uncontrolled lateral object translations, we introduced an object displacement vector (Fig. 3.1b), $\boldsymbol{D} = (D_x, D_y)$, that has two components, defined as independent, uniformly distributed random variables:

$$D_x \sim U(-\Delta_x, \Delta_x)$$
$$D_y \sim U(-\Delta_y, \Delta_y) \tag{3.9}$$

The standard diffractive network model (shown in Fig. 3.2a) was trained (*tr*) with $\Delta_x = \Delta_y = \Delta_{tr} = 0$, and was then tested under different levels of input object position shifts by sweeping the values of $\Delta_x = \Delta_y = \Delta_{test}$ from 0 to 33.92λ with steps of 0.53λ. Stated differently, the final test accuracy corresponding to each Δ_{test} value reflects the image classification performance of the same diffractive network model that was tested with 10,000 different object positions randomly chosen within the range set by Δ_{test} (see Fig. 3.3a for exemplary test objects). This analysis revealed that the blind inference accuracy of the standard D^2NN design ($\Delta_{tr} = 0$) which achieves 97.64% under $\Delta_{test} = 0$ quickly falls below 90% as the input objects starts to move within the range $\mp$3.5λ (blue curve in Fig. 3.3, defined with $\Delta_{tr} = 0$). As the area covered by the possible object shifts is increased further, the inference accuracy of this native network model decreases rapidly (see Fig. 3.3).

In this conventional design approach, the optical forward model of the diffractive network training assumes that the input objects inside the sample FOV are free-from any type of undesired geometrical variations, i.e., $\Delta_{tr} = 0$. Hence, the diffractive layers are *not* challenged to process optical waves coming from input objects at different spatial locations, possibly

overfitting to the assumed FOV location. As a result, the inference performance of the resulting diffractive network model becomes dependent on the relative lateral location of the input object with respect to the plane of the diffractive layers and the output detectors.

To mitigate this problem, we adopted a training strategy inspired by data augmentation techniques used in deep learning. According to this scheme, each training image sample in a batch is randomly shifted, based on a realization of the displacement vector ($\boldsymbol{D}$), and subsequently, the loss function is computed by propagating these randomly shifted object fields through the diffractive network (see the Methods for details). Using this training scheme, we designed 5 different diffractive network models based on different ranges of object displacement, i.e., $\Delta_x = \Delta_y = \Delta_{tr} = 2.12\lambda, 4.24\lambda, 8.48\lambda, 16.96\lambda \; and \; 33.92\lambda$ (see Eq. 3.1). Figure 3.3 illustrates the MNIST image classification accuracies provided by these 5 new diffractive network models as a function of Δ_{test}. Comparison between the diffractive network models trained with $\Delta_{tr} = 0$ (blue) and $\Delta_{tr} = 2.12\lambda$ (red) reveals that due to the data augmentation introduced by the small object shifts during the training, the latter can achieve an improved inference accuracy of 98.00% for MNIST digits under $\Delta_{test} = 0$. Furthermore, the diffractive network trained with $\Delta_{tr} = 2.12\lambda$ can maintain its classification performance when the input objects are randomly shifted within a certain lateral range (see the right shift of the red curve in Fig. 3.3). Similarly, training a diffractive network model with $\Delta_{tr} = 4.24\lambda$ (yellow curve in Fig. 3.3) also results in a better classification accuracy of 97.75% when compared to the 97.64% achieved by the standard model ($\Delta_{tr} = 0$) under $\Delta_{test} = 0$. In addition, this new diffractive model exhibits further resilience to random shifts of the objects within the input FOV, which is indicated by the stronger right shift of the yellow curve in Fig. 3. For example, for $\Delta_{test} = 3.71\lambda$ in Fig. 3.3, the input test objects are randomly shifted in x and y by an amount determined by $D_x \sim U(-3.71\lambda, 3.71\lambda)$ and

$D_y \sim U(-3.71\lambda, 3.71\lambda)$, respectively, and this results in a classification accuracy of 97.07% for the new diffractive model ($\Delta_{tr} = 4.24\lambda$), whereas the inference accuracy of the standard model ($\Delta_{tr} = 0$) decreases to 89.88% under the same random lateral shifts of the input test objects.

Further increasing the range of the object location uncertainty, e.g., to $\Delta_{tr} = 8.48\lambda$ (purple curve in Fig. 3.3), we start to observe a trade-off between the peak inference accuracy and the resilience of the diffractive network to random object shifts. For instance, the diffractive optical network trained with $\Delta_{tr} = 8.48\lambda$ can achieve a peak classification accuracy of 95.55%, which represents a ~2% accuracy compromise with respect to the native diffractive network model ($\Delta_{tr} = 0$) tested under $\Delta_{test} = 0$. However, using such a large object location uncertainty in the training phase also results in a rather flat accuracy curve over a much larger Δ_{test} range as shown in Fig. 3.3; in other words, this design strategy expands the effective input object FOV that can be utilized for the desired machine learning task. For example, if the test objects were to freely move within the area defined by $\Delta_x = \Delta_y = \Delta_{test} = 6.89\lambda$, the diffractive network model trained with $\Delta_{tr} = 8.48\lambda$ (purple curve in Fig. 3.3) brings a >30% inference accuracy advantage compared to the standard model (blue curve in Fig. 3.3). The resulting layer thickness profiles for this diffractive optical network design trained with $\Delta_{tr} = 8.48\lambda$ are also shown in Fig. 3.2b.

For the case where Δ_{tr} was set to be 16.96λ, the mean test classification accuracy over the range $0 < \Delta_{test} < \Delta_{tr}$ is observed to be 90.46% (see the green curve in Fig. 3.3b). The relatively more pronounced performance trade-off in this case can be explained based on the increased input FOV. Stated differently, with larger Δ_{tr} values, the effective input FOV of the diffractive network is increased, and the dimensionality of the solution space[100] provided by a diffractive network design with a limited number of layers (and neurons) might not be sufficient

to provide the desired solution when compared to a smaller input FOV diffractive network design. The use of wider diffractive layers (i.e., larger number of neurons per layer) can be a strategy to further boost the inference accuracy over larger Δ_{tr} values (or larger effective input FOVs), which will be further discussed and demonstrated later in our analysis below (see Fig. 3.4b).

As an alternative design strategy, the detector plane configuration shown in Fig. 3.1a can also be replaced with a differential detection scheme[79] to mitigate this relative drop in blind inference accuracy for designs with large Δ_{tr}. In this scheme, instead of assigning a single optoelectronic detector per class, we designate two detectors to each data class and represent the corresponding class scores based on the normalized difference between the optical signals collected by each detector pair. Figure 3.4a illustrates a comparison between the blind classification accuracies of standard (solid curves) and differential (dashed curves) diffractive network designs, when they were trained with random lateral shifts of the input objects. For all of these designs, except the $\Delta_{tr} = 33.92\lambda$ case, the differential diffractive networks achieve higher classification accuracies throughout the entire testing range, showing their superior robustness and adaptability to input field variations compared to their non-differential counterparts. For example, the peak inference accuracy (95.55%) achieved by the diffractive optical network trained with $\Delta_{tr} = 8.48\lambda$ (solid purple curve in Fig. 3.4a) increases to 97.33% using the differential detection scheme (dashed purple curve in Fig. 3.4a). As another example, for $\Delta_{tr} = 16.96\lambda$, the mean classification accuracy of the differential diffractive network over $0 < \Delta_{test} < \Delta_{tr}$ yields 93.38%, which is ~3% higher compared to the performance of its non-differential counterpart for the same test range.

On the other hand, enlarging the uncertainty in the input object translation further, e.g., $\Delta_{tr} = 33.92\lambda$, starts to balance out the benefits of using differential detection at the output plane (see the solid and dashed blue curves in Fig. 3.4, which closely follow each other). In fact, when Δ_x and Δ_y in Eq. 3.1 are large enough, such as $\Delta_{tr} = 33.92\lambda$, the effective input FOV increases considerably with respect to the size of the diffractive layers; as we discussed earlier, the use of wider diffractive layers with larger numbers of neurons per layer could be used to mitigate this and improve inference performance of D^2NN designs that are trained with relatively large Δ_{tr} values. To shed more light on this, using $\Delta_{tr} = 33.92\lambda$ we trained two additional diffractive optical network models with wider diffractive layers that cover m=4 and m=9 fold larger number of neurons per layer compared to the standard design (m=1) that has 40K neurons per diffractive layer; stated differently, each diffractive layer of these two new designs contain (2×200)×(2×200) = 4×40K and (3×200)×(3×200) = 9×40K neurons per layer, covering 5 diffractive layers, same as the standard D^2NN design. The comparison of the blind classification accuracies of these 5-layer D^2NN designs with m=1, 4 and 9, all trained with $\Delta_{tr} = 33.92\lambda$, reveals that an increase in the width of the diffractive layers not only increases the input numerical aperture (NA) of the diffractive network, but also significantly improves the classification accuracies even under large Δ_{test} (see Fig. 3.4b). For example, the D^2NN design with $\Delta_{tr} = 33.92\lambda$ and m=4 achieves classification accuracies of 83.08% and 85.76% for the testing conditions, $\Delta_{test} = 0.0\lambda$ and $\Delta_{test} = \Delta_{tr} = 33.92\lambda$, respectively. With the same Δ_{test} values, the diffractive network with m=1, i.e., 40K neurons per layer can only achieve classification accuracies of 79.23% and 81.98%, respectively. The expansion of the diffractive layers to accommodate 9×40K neurons per layer (m=9), further increases the mean classification accuracies over the entire Δ_{test} range, as illustrated in Fig. 3.4b.

Next, we expanded the presented training approach to design diffractive optical network models that are resilient to the *scale* of the input objects. To this end, similar to Eqs. 3.1a and 3.1b, we defined a scaling parameter, $K \sim U(1-\zeta, 1+\zeta)$, randomly covering the scale range $(1-\zeta, 1+\zeta)$ determined by the hyperparameter, ζ. According to this formulation, for a given value of K, the physical size of the input object is scaled up ($K > 1$) or down ($K < 1$); see Fig. 3.5a. Based on this formulation, in addition to the standard D^2NN design with $\zeta_{tr} = 0$, we trained 4 new diffractive network models with ζ_{tr} = 0.1, 0.2, 0.4 and 0.8. The resulting diffractive network models were then tested by sweeping ζ_{test} from 0 to 0.8 with steps of 0.02 and for each case, the classification accuracy on testing data attained by each diffractive model was computed (see Fig. 3.5b). This analysis reveals that the resulting diffractive network designs are rather resilient to random scaling of the input objects, maintaining a competitive inference performance over a large range of object shrinkage or expansion (Fig. 3.5b). Similar to the case shown in Fig. 3.3, the relatively small values of ζ_{tr}, e.g., 0.1 (red curve in Fig. 3.5b) or 0.2 (yellow curve in Fig. 3.5b), effectively serve as data augmentation and the corresponding diffractive network models achieve higher peak inference accuracies of 97.84% ($\zeta_{tr} = 0.1$) and 97.88% ($\zeta_{tr} = 0.2$) compared to the 97.64% achieved by the standard design ($\zeta_{tr} = 0$). Furthermore, the comparison between the shift- and scale-invariant diffractive optical network models trained with $\Delta_{tr} = 16.96\lambda$ (green curve in Fig. 3.3b) and $\zeta_{tr} = 0.8$ (green curve in Fig. 3.5b) is highly interesting since the effective FOVs induced by these two training parameters at the input/object plane are quite comparable, resulting in ~1.87× and 1.8× of the FOV of the standard design ($\Delta_{tr} = \zeta_{tr} = 0$), respectively. Despite these comparable effective FOVs at the input plane, the diffractive network trained against random scaling, $\zeta_{tr} = 0.8$, achieves nearly ~6% higher inference accuracy compared to the shift-invariant design, $\Delta_{tr} = 16.96\lambda$. The mean

classification accuracy provided by this scale-invariant diffractive optical network model ($\zeta_{tr} = 0.8$) over the entire testing range, $0 < \zeta_{test} < 0.8$, is found to be 96.57% (Fig. 3.5b), which is only ~1% lower than that of the standard diffractive design tested in the absence of random object scaling ($\zeta_{test} = 0$). The difference in adaptation capability of diffractive optical networks against random translation and scaling of input objects can be attributed to the changes in the effective space-bandwidth product at the input plane induced by these two transformations. According to our scaling model, differently scaled versions of the same MNIST digits has identical space-bandwidth products, since the larger object also has larger features vice versa, preserving the total information content. On the other hand, translation operation does not affect the size of local object features, thus preserves the spatial frequency bandwidth. Consequently, every possible object location in space expands the total space-bandwidth product at the input plane of the subsequent diffractive network, contributing to the difficulty of the inference task at hand in a more significant way.

To explore if there is a large performance gap between the classification accuracies attained for de-magnified and magnified input objects, next we *separately* tested the diffractive optical network models in Fig. 3.5b for the case of expansion-only, i.e., $K \sim U(1, 1+\zeta)$ and shrinkage-only, i.e., $K \sim U(1-\zeta, 1)$; see Fig.3. 5c. A comparison of the solid (expansion-only) and the dashed (shrinkage-only) curves in Fig. 3.5c reveals that, in general, diffractive networks' resilience toward object expansion and object shrinkage is similar. For instance, for the case of $\zeta_{tr} = 0.4$ (purple curves in Fig. 3.5c) the mean classification accuracy difference observed between the expansion-only vs. shrinkage-only testing is only 0.04% up to the point that the testing range is equal to that of the training, i.e., $\zeta_{test} = \zeta_{tr}$. Similarly, for ζ_{tr}= 0.8 the mean classification accuracy difference observed between the expansion-only vs. shrinkage-only

testing is ~0.75%. When analyzing these results reported in Fig. 3.5c, one should carefully consider the fact that for a fixed choice of ζ parameter there is an inherent asymmetry in expansion and shrinkage percentages; for example, for $\zeta_{test} = 0.8$, K can take values in the range (0.2,1.8), where the extreme cases of 0.2 and 1.8 correspond to 5× shrinkage and 1.8× expansion of the input object, respectively. Therefore, the curves reported in Fig. 3.5c for expansion-only vs. shrinkage-only testing naturally contain different percentages of scaling with respect to the original size of the input objects.

Next, we expanded the presented framework to handle input object *rotations*. Figure 3.6 illustrates an equivalent analysis as in Fig. 3.3, except that the input objects are now rotating, instead of shifting, around the optical axis, according to a uniformly distributed random rotation angle, $\Theta \sim U(-\theta, \theta)$, where $\Theta < 0$ and $\Theta > 0$ correspond to clockwise and counterclockwise rotation as depicted in Fig. 3.1b, respectively. In this comparative analysis, six different diffractive network models trained with θ_{tr} values taken as 0° (standard design), 5°, 10°, 20°, 30° and 60° were tested as a function of θ_{test} taking values between 0° and 60° with a step size of 1°, i.e., $\Theta \sim U(-\theta_{test}, \theta_{test})$. Similar to the case of scale-invariant designs reported in Fig. 3.5, these diffractive network models trained with different θ_{tr} values can build up strong resilience against random object rotations, almost without a compromise in their inference. In fact, training with $\theta_{tr} \leq 20°$ (red, yellow and purple curves in Fig. 3.6b) improves the peak inference accuracy over the standard design ($\theta_{tr} = 0°$). When $\theta_{tr} = 30°$ (green curve in Fig. 3.6b), the inference of the diffractive optical network is relatively flat as a function of θ_{test}, achieving a classification accuracy of 97.51% and 96.68% for $\theta_{test} = 0^0$ and $\theta_{test} = 30^0$, respectively, clearly demonstrating the advantages of the presented design framework.

Finally, we investigated the design of diffractive optical network models that were trained to simultaneously accommodate two of the three commonly encountered input objects transformations, i.e., random lateral shifting, scaling and in-plane rotation. Table 3.1 reports the resulting classification accuracies of these newly trained D^2NN models, where the inference performance of the corresponding diffractive optical network was tested with the same level of random object transformation as in the training, i.e., $\Delta_{tr} = \Delta_{test}$, $\zeta_{tr} = \zeta_{test}$, $\theta_{tr} = \theta_{test}$. The results in Table 3.1 reveal that these diffractive network designs can maintain their inference accuracies over 90%, building up resilience against unwanted, yet practically-inevitable object transformations and variations. The thickness profile of the diffractive layers constituting the D^2NN designs trained with the object transformation parameter pairs: ($\Delta_{tr} = 2.12\lambda, \theta_{tr} = 10°$), ($\Delta_{tr} = 2.12\lambda, \zeta_{tr} = 0.4$) and ($\theta_{tr} = 10°, \zeta_{tr} = 0.4$) reported in Table 3.1 are illustrated in Fig. 3.7. The confusion matrices provided by these three diffractive network models computed under $\Delta_{tr} = \Delta_{test}$, $\zeta_{tr} = \zeta_{test}$, and $\theta_{tr} = \theta_{test}$, are also reported in Fig. 3.8.

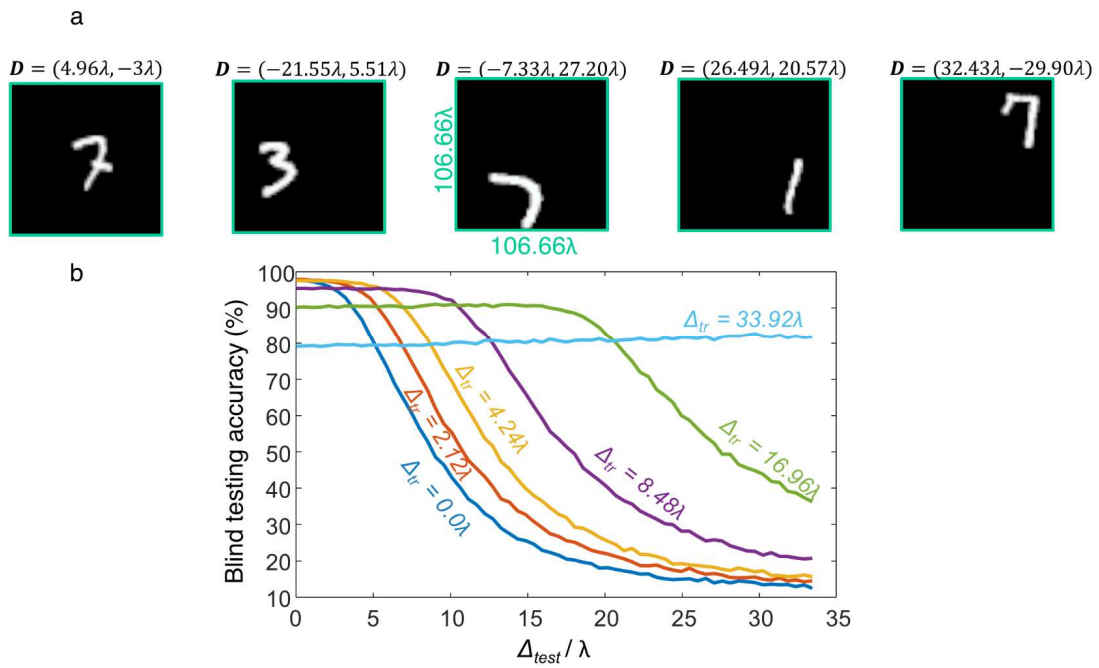

Fig. 3.3 Shift-invariant diffractive optical networks. (a) Randomly shifted object samples from the MNIST test dataset. Green frame around each object demonstrates the size of the diffractive layers (106.66λ×106.66λ). (b)The blind inference accuracies provided by six different diffractive network models trained with $\Delta_x = \Delta_y = \Delta_{tr}$, taken as 0.0λ (blue), 2.12λ (red), 4.24λ (yellow), 8.48λ (purple), 16.96λ (green), 33.92λ (light-blue) when they were tested under different levels random object shifts with the control parameter, $\Delta_x = \Delta_y = \Delta_{test}$, swept from 0.0λ to 33.92λ.

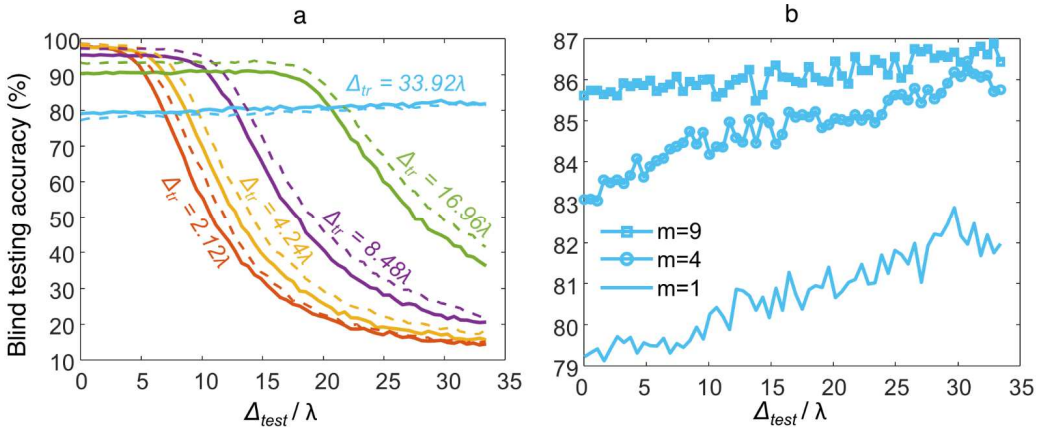

Fig. 3.4 Different design strategies that can improve the performance of shift-invariant diffractive optical networks. (a) The comparison between the inference accuracies of standard (solid curves) and differential (dashed curves) diffractive optical networks trained using various Δ_{tr} values. (b) Blind testing classification accuracies of three non-differential, 5-layer D^2NN designs that have m×40K optical neurons per layer, with m=1, 4 and 9. All these diffractive optical networks were trained using Δ_{tr}= 33.92λ. The diffractive network designs with wider diffractive layers and more neurons per layer can generalize more effectively to random object translations.

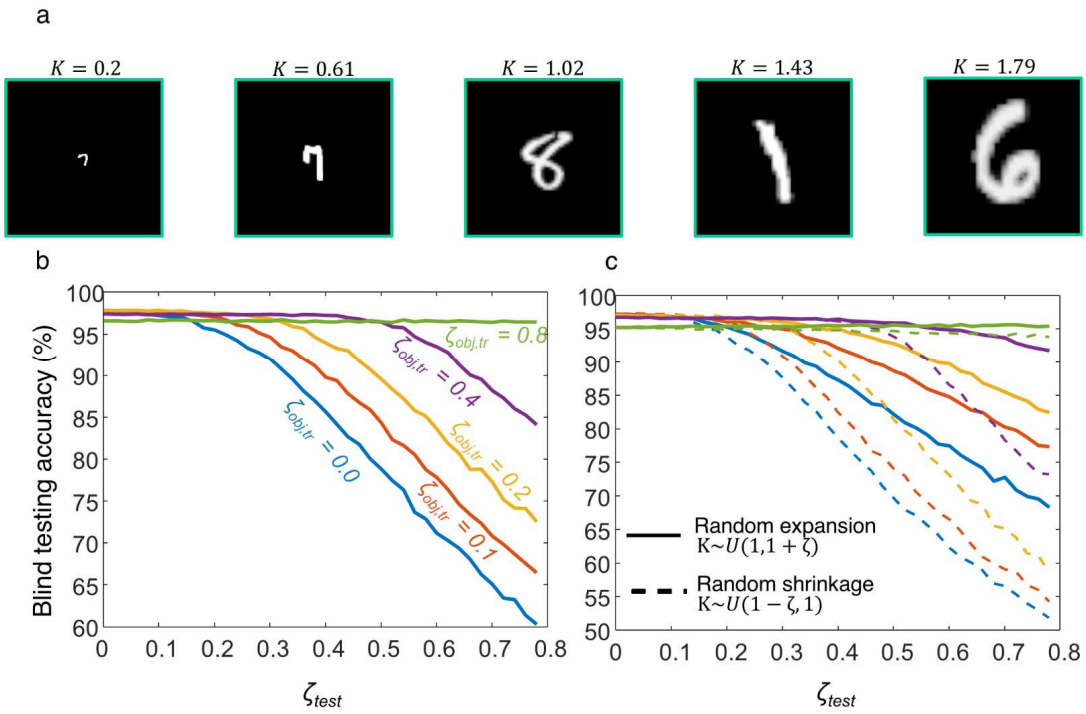

Fig. 3.5 Scale-invariant diffractive optical networks. (a) Randomly scaled object examples from the MNIST test dataset. Green frame around each object demonstrates the size of the diffractive layers. (b) The blind inference accuracies provided by five different D^{2}NN models trained with $\zeta = \zeta_{tr}$, taken as 0.0 (blue), 0.1 (red), 0.2 (yellow), 0.4 (purple) and 0.8 (green); the resulting models were tested under different levels random object scaling with the parameter, $\zeta = \zeta_{test}$, swept from 0.0 to 0.8. (c) The classification performance of the diffractive networks in (b) for the case of expansion-only (solid curves) and shrinkage-only (dashed curves).

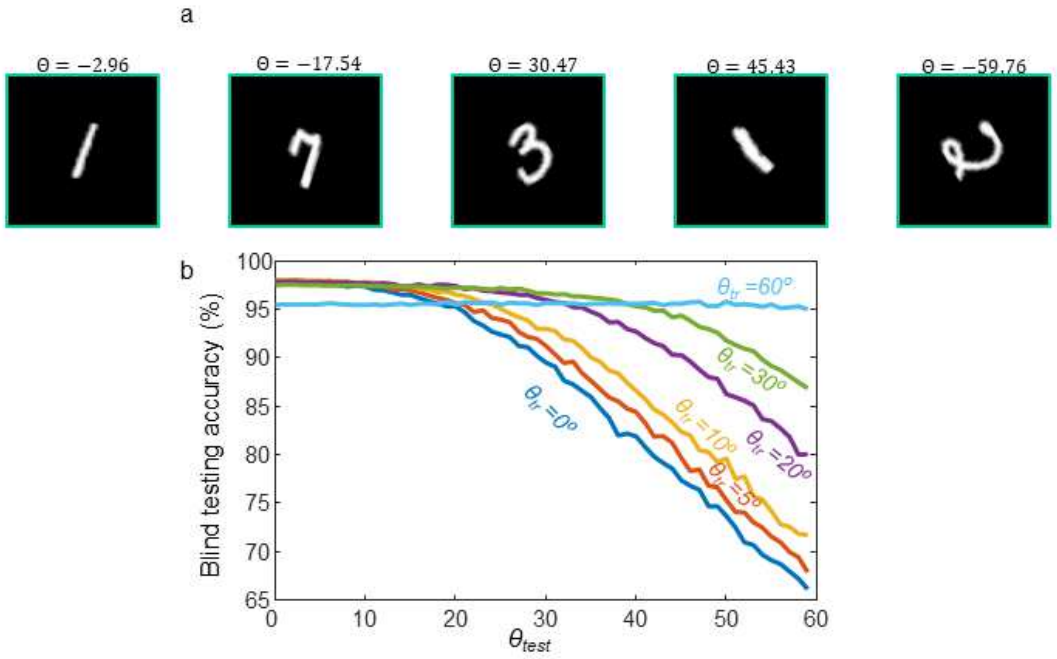

Fig. 3.6 Rotation-invariant diffractive optical networks. (a) Randomly rotated object examples from the MNIST test dataset. Green frame around each object demonstrates the size of the diffractive layers. (b) The blind inference accuracies provided by five different diffractive network models trained with $\theta = \theta_{tr}$, taken as 0° (blue), 5° (red), 10° (yellow), 20° (purple), 30° (green) and 60° (light-blue) when they were tested under different levels of random object rotations with the parameter, $\theta = \theta_{test}$, swept from 0° to 60°, covering both clockwise and counter-clockwise image rotations.

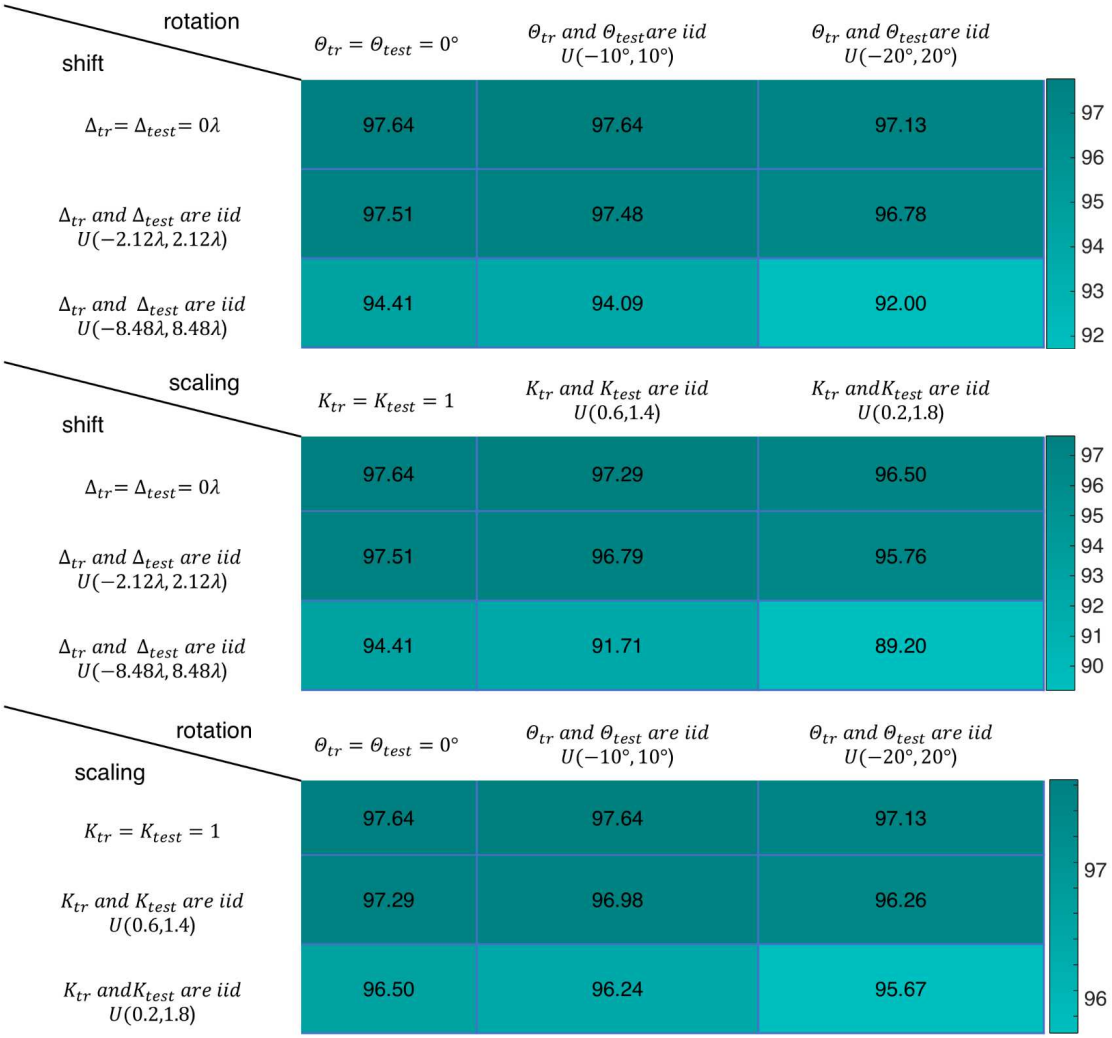

Table. 3.1 The blind inference accuracy of the D^{2}NN models trained against the combinations of the three object field transformations investigated in this work: (upper) shift-rotation, (middle) shift-scaling, (lower) rotation-scaling.

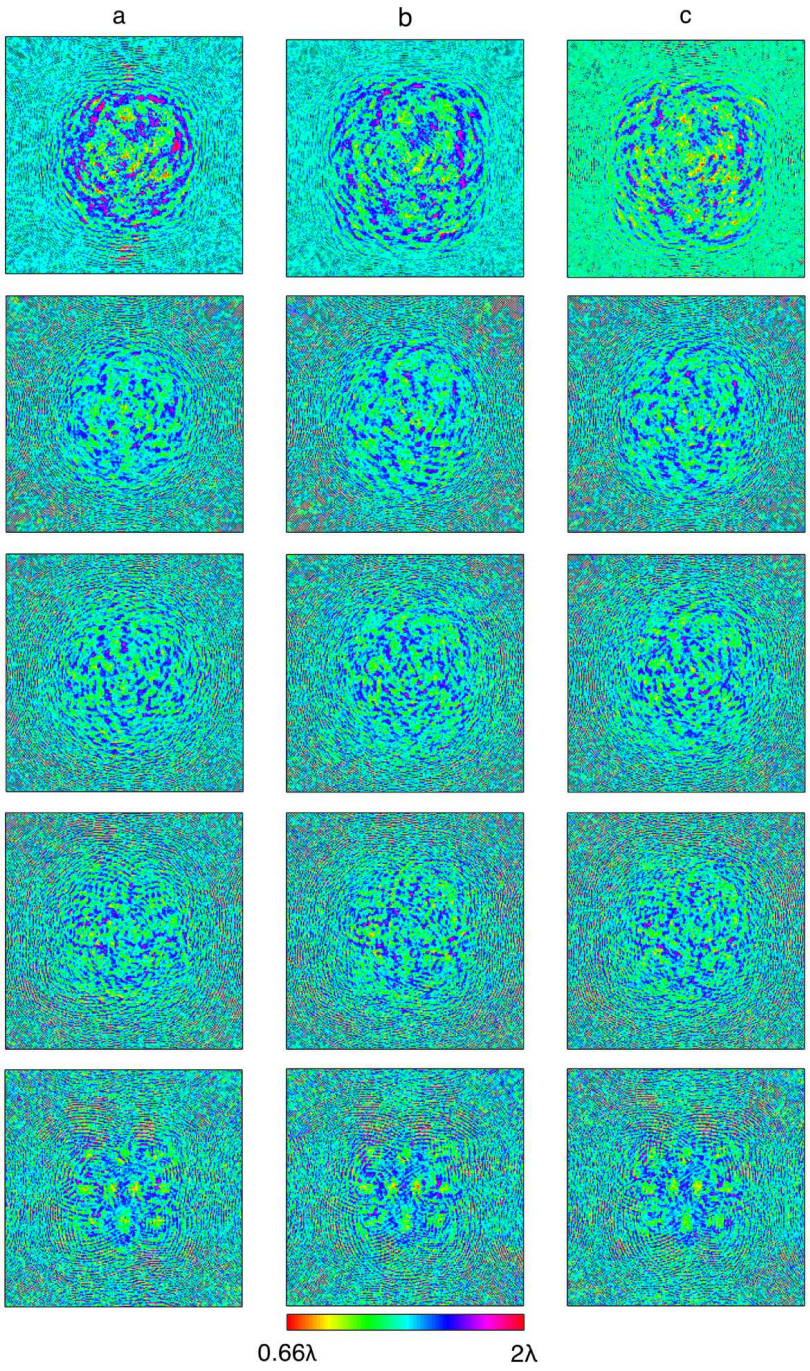

Fig. 3.10 The thickness profiles of the diffractive networks reported in Table 3.1. (a) $\Delta_{tr} = 2.12\lambda$, $\theta_{tr} = 10°$; (b) $\Delta_{tr} = 2.12\lambda$, $\zeta_{tr} = 0.4$; (c) $\theta_{tr} = 10°$, $\zeta_{tr} = 0.4$.

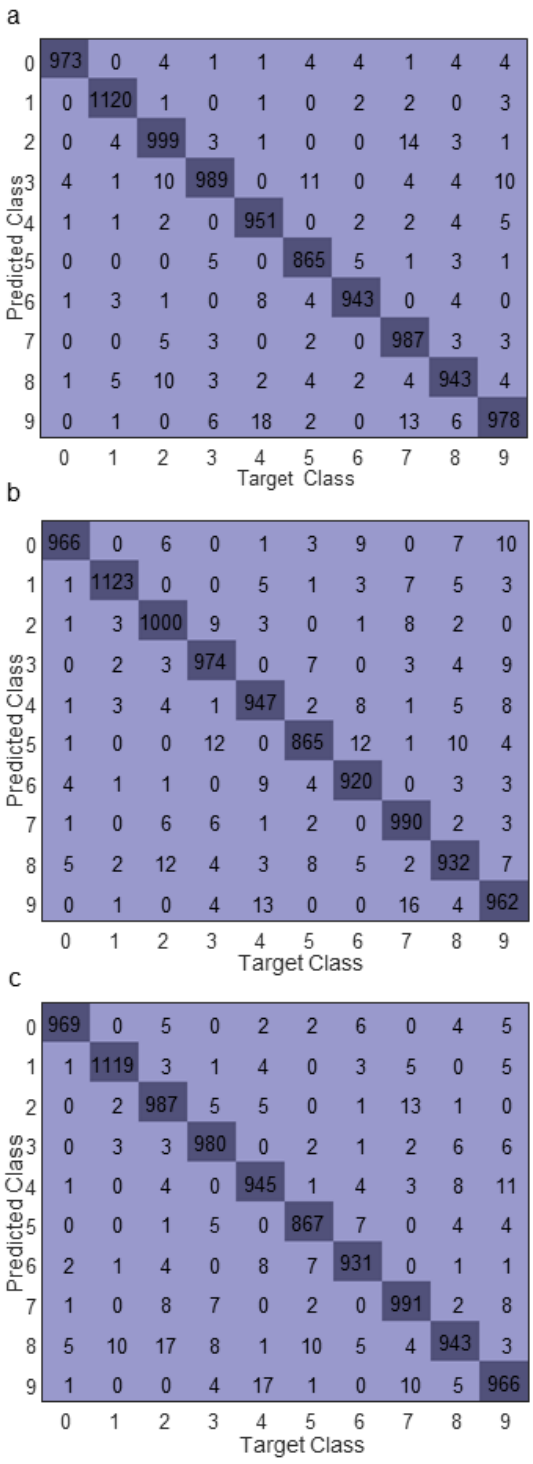

Fig. 3.11 The confusion matrices achieved by the diffractive network designs shown in Fig. S1. (a) $\Delta_{tr} = \Delta_{test} = 2.12\lambda$, $\theta_{tr} = \theta_{test} = 10°$; (b) $\Delta_{tr} = \Delta_{test} = 2.12\lambda$, $\zeta_{tr} = \zeta_{test} = 0.4$; (c) $\theta_{tr} = \theta_{test} = 10°$, $\zeta_{tr} = \zeta_{test} = 0.4$.

3.3 Methods

D^{2}NN framework formulates the all-optical object classification problem from the point-of-view of training the physical features of matter inside a diffractive optical black-box. In this study, we modeled each D^{2}NN using 5 successive modulation layers, each representing a two-dimensional, thin modulation component (Fig. 3.1a). The optical modulation function of each diffractive layer was sampled with a period of 0.53λ over a regular 2D grid of coordinates, with each point representing the transmittance coefficient of a diffractive feature, i.e., an optical "neuron". Following earlier work[80,99,101,103], we selected the material thickness, h, as the trainable physical parameter of each neuron,

$$h = Q_4\left(\frac{\sin(h_a)+1}{2}(h_m - h_b)\right) + h_b \qquad (3.2),$$

According to Eq. 3.2, the material thickness over each diffractive neuron is defined as a function of an auxiliary variable, h_a. The function, $Q_n(.)$, represents the n-bit quantization operator and h_m, h_b denote the pre-determined hyperparameters of our forward model determining the allowable range of thickness values, $[h_b, h_m]$. The thickness in Eq. 3.2 is related to the transmittance coefficient of the corresponding diffractive neuron through the complex-valued refractive index (τ) of the optical material used to fabricate the resulting D^{2}NN, i.e., $\tau(\lambda) = n(\lambda) + j\kappa(\lambda)$, with λ denoting the wavelength of the illumination light. Based on this, we can express the transmission coefficient, $t(x_q, y_p, z_k)$, of a diffractive neuron located at (x_q, y_p, z_k) as;

$$t(x_q, y_p, z_k) = \exp\left(-\frac{2\pi\kappa h_{q,p}^k}{\lambda}\right)\exp\left(j(n-n_s)\frac{2\pi h_{q,p}^k}{\lambda}\right) \qquad (3.3),$$

where $h_{q,p}^{k}$ refers to the material thickness over the corresponding neuron computed using Eq. 3.2, and n_s is the refractive index of the medium, surrounding the diffractive layers; without loss of generality, we assumed $n_s = 1$ (air). Based on the earlier demonstrations of diffractive optical networks[77,80,99,101,103], we assumed the optical modulation surfaces in our diffractive optical networks are made of a material with τ = 1.7227 + j0.031. Accordingly, the h_m and h_b were selected as 2λ and 0.66λ, respectively, as illustrated in Fig. 3.2 and Fig. 3.7.

The 2D complex modulation function, $T(x, y, z_k)$, of a diffractive surface, S_k, located at $z = z_k$, can be written as:

$$T(x, y, z_k) = \sum_q \sum_p t(x_q, y_p, z_k)\, P(x - qw_x, y - pw_y, z_k) \tag{3.4},$$

where the w_x and w_y denote the width of a diffractive neuron in x and y directions, respectively (both taken as 0.53λ). $P(x, y, z_k)$ represents the 2D interpolation kernel which we assumed to be an ideal rectangular function in the following form,

$$P(x, y, z_k) = \begin{cases} 1, & |x| < \left(\frac{w_x}{2}\right) and\ |y| < \left(\frac{w_y}{2}\right) \\ 0, & otherwise \end{cases} \tag{3.5}.$$

The light propagation in the presented diffractive optical networks were modeled based on the digital implementation of the Rayleigh-Sommerfeld diffraction equation, using an impulse response defined as:

$$w(x, y, z) = \frac{z}{r^2}\left(\frac{1}{2\pi r} + \frac{1}{j\lambda}\right)\exp\left(\frac{j2\pi r}{\lambda}\right) \tag{3.6},$$

where $r = \sqrt{x^2 + y^2 + z^2}$. Based on this, the wave field synthesized over a surface at $z = z_{k+1}$, $U(x, y, z_{k+1})$, by a trainable diffractive layer, S_k, located at $z = z_k$, can expressed as;

$$U(x, y, z_{k+1}) = U'(x, y, z_k) * w(x, y, z_{k+1} - z_k) \quad (3.7),$$

where $U'(x, y, z_k) = U(x, y, z_k)T(x, y, z_k)$ is the complex wave field immediately after the diffractive layer, k, and $*$ denotes the 2D convolution operation. In this optical forward model, the layer-to-layer distances were taken as 40λ for the diffractive network architectures that has 40K neurons on each layer to induce connections between all the neurons of two successive layers based on Eq. 3.6. For the diffractive network architectures constituting, m=4 and m=9, times larger diffractive layers as depicted in Fig. 3.4b, the layer-to-layer distances were set to be $(m)^{0.5} \times 40\lambda$ preserving the diffraction cone angle of optical connections between the successive layers of these network models for a fair comparison. Therefore, the improvement in inference accuracy for randomly shifting objects demonstrated in Fig. 3.4b, comes at the expense of using larger diffractive layers separated with larger distances increasing the cost and both the lateral and axial size of the diffractive network.

Based on the above outlined optical forward model, if we let the complex-valued object transmittance, $T(x, y, z_0)$, over the input FOV be located at a surface defined with $k = 0$, then the complex field and the associated optical intensity distribution at the *output/detector plane* of a 5-layer diffractive optical network architecture shown in Fig. 3.1a, can be expressed as $U(x, y, z_6)$ and $I = |U(x, y, z_6)|^2$, respectively. In our forward training model, we assumed that each class detector collects an optical signal, Γ_c, that is computed through the integration of the output intensity, I, over the corresponding detector active area (6.4λ×6.4λ per detector). For a given dataset with C classes, the standard D^2NN architecture in Fig. 3.1a employs C detectors at

the output plane, each representing a data class; C=10 for MNIST dataset. Accordingly, at each training iteration, after the propagation of the input object to the output plane (based on Eqs. 3.6 and 3.7), a vector of optical signals, $\mathbf{\Gamma}$, is formed and then normalized to get $\mathbf{\Gamma}'$ using the following relationship:

$$\mathbf{\Gamma}' = \frac{\mathbf{\Gamma}}{\max\{\mathbf{\Gamma}\}} \times T_s \tag{3.8},$$

where T_s is a constant temperature parameter[104,105]. Next, the class score of the c[th] data class, σ_c, is computed as:

$$\sigma_c = \frac{\exp(\Gamma_c')}{\sum_{c\,\epsilon\,C} \exp(\Gamma_c')} \tag{3.9}.$$

In Eq. 3.9, Γ_c' denotes the normalized optical signal collected by the detector, c, computed as in Eq. 3.8. At the final step, the classification loss function, $\mathcal{L}$, in the form of the cross-entropy loss defined in Eq. 3.10 is computed for the subsequent error-backpropagation and update of the diffractive layers:

$$\mathcal{L} = -\sum_{c\,\epsilon\,C} g_c \log(\sigma_c) \tag{3.10},$$

where $\boldsymbol{g}$ denotes the one-hot ground truth label vector.

For the digital implementation of the diffractive optical network training outlined above, we developed a custom-written code in Python (v3.6.5) and TensorFlow (v1.15.0, Google Inc.). The backpropagation updates were calculated using the Adam[106] optimizer with its parameters set to be the default values as defined by TensorFlow and kept identical in each model. The learning rate was set to be 0.001 for all the diffractive network models presented here. The training batch

sizes were taken as 50 and 20 for the diffractive network designs with 40K neurons per layer and wider diffractive networks reported in Fig. 3.4b, respectively. The training of a 5-layer diffractive optical network with 40K diffractive neurons per layer takes ~6 hours using a computer with a GeForce GTX 1080 Ti Graphical Processing Unit (GPU, Nvidia Inc.) and Intel® Core ™ i7-8700 Central Processing Unit (CPU, Intel Inc.) with 64 GB of RAM, running Windows 10 operating system (Microsoft). The training of a wider diffractive network presented in Fig. 3.4b, on the other hand, takes ~30 hours based on the same system configuration due to the larger light propagation windows used in the forward optical model. Since the investigated object transformations were implemented through a custom-developed bilinear interpolation code written based on TensorFlow functions, it only takes ~50 sec longer to complete an epoch with the presented scheme compared to the standard training of D^2NNs.

Chapter 4 All-optical Information Processing Capacity of Diffractive Surfaces

Parts of this chapter have previously been published in O. Kulce et al. “All-optical Information Processing Capacity of Diffractive Surfaces”, Light Science & Applications, DOI:. 10.1038/s41377-020-00439-9. This chapter presents in depth analysis on the multi-layer diffractive free-space optical processors and defines the upper bound on their information processing capacity.

The precise engineering of materials and surfaces has been at the heart of some of the recent advances in optics and photonics. These advances related to the engineering of materials with new functionalities have also opened up exciting avenues for designing trainable surfaces that can perform computation and machine learning tasks through light-matter interactions and diffraction. Here, we analyse the information processing capacity of coherent optical networks formed by diffractive surfaces that are trained to perform an all-optical computational task between a given input and output field-of-view. We show that the dimensionality of the all-optical solution space covering the complex-valued transformations between the input and output fields-of-view is linearly proportional to the number of diffractive surfaces within the optical network, up to a limit that is dictated by the extent of the input and output fields-of-view. Deeper diffractive networks that are composed of larger numbers of trainable surfaces can cover a higher-dimensional subspace of the complex-valued linear transformations between a larger input field-of-view and a larger output field-of-view and exhibit depth advantages in terms of their statistical inference, learning and generalization capabilities for different image classification tasks when compared with a single trainable diffractive surface. These analyses

and conclusions are broadly applicable to various forms of diffractive surfaces, including, e.g., plasmonic and/or dielectric-based metasurfaces and flat optics, which can be used to form all-optical processors.

4.1 Introduction

The ever-growing area of engineered materials has empowered the design of novel components and devices that can interact with and harness electromagnetic waves in unprecedented and unique ways, offering various new functionalities [107–120]. Owing to the precise control of material structure and properties as well as the associated light-matter interaction at different scales, these engineered material systems, including, e.g., plasmonics, metamaterials/metasurfaces and flat optics, have led to fundamentally new capabilities in the imaging and sensing fields, among others [121–130]. Optical computing and information processing constitute yet another area that has harnessed engineered light-matter interactions to perform computational tasks using wave optics and the propagation of light through specially devised materials[75,92,131,73,69,74,84,70,72,20,22,71,81,77]. These approaches and many others highlight the emerging uses of trained materials and surfaces as the workhorse of optical computation.

Here, we investigate the information processing capacity of trainable diffractive surfaces to shed light on their computational power and limits. An all-optical diffractive network is physically formed by a number of diffractive layers/surfaces and the free-space propagation between them (see Fig. 4.1a). Individual transmission and/or reflection coefficients (i.e., neurons) of diffractive surfaces are adjusted or trained to perform a desired input-output transformation task as the light diffracts through these layers. Trained with deep-learning-based error back-propagation methods, these diffractive networks have been shown to perform machine learning tasks such as image classification and deterministic optical tasks including, e.g., wavelength demultiplexing, pulse shaping and imaging[77–80,99,101,132].

The forward model of a diffractive optical network can be mathematically formulated as a complex-valued matrix operator that multiplies an input field vector to create an output field vector at the detector plane/aperture. This operator is designed/trained using, e.g., deep learning to transform a set of complex fields (forming, e.g., the input data classes) at the input aperture of the optical network into another set of corresponding fields at the output aperture (forming, e.g., the data classification signals) and is physically created through the interaction of the input light with the designed diffractive surfaces as well as free-space propagation within the network (Fig. 4.1a).

In this paper, we investigate the dimensionality of the all-optical solution space that is covered by a diffractive network design as a function of the number of diffractive surfaces, the number of neurons per surface, and the size of the input and output fields-of-view. With our theoretical and numerical analysis, we show that the dimensionality of the transformation solution space that can be accessed through the task-specific design of a diffractive network is linearly proportional to the number of diffractive surfaces, up to a limit that is governed by the extent of the input and output fields-of-view. Stated differently, adding new diffractive surfaces into a given network design increases the dimensionality of the solution space that can be all-optically processed by the diffractive network until it reaches the linear transformation capacity dictated by the input and output apertures (Fig. 4.1a). Beyond this limit, the addition of new trainable diffractive surfaces into the optical network can cover a higher-dimensional solution space over larger input and output fields-of-view, extending the space-bandwidth product of the all-optical processor.

Our theoretical analysis further reveals that, in addition to increasing the number of diffractive surfaces within a network, another strategy to increase the all-optical processing capacity of a diffractive network is to increase the number of trainable neurons per diffractive surface. However, our numerical analysis involving different image classification tasks demonstrates that this strategy of creating a higher-numerical-aperture (NA) optical network for all-optical processing of the input information is not as effective as increasing the number of diffractive surfaces in terms of the blind inference and generalization performance of the network. Overall, our theoretical and numerical analyses support each other, revealing that deeper diffractive networks with larger numbers of trainable diffractive surfaces exhibit depth advantages in terms of their statistical inference and learning capabilities compared with a single trainable diffractive surface.

The presented analyses and conclusions are generally applicable to the design and investigation of various coherent all-optical processors formed by diffractive surfaces such as, e.g., metamaterials, plasmonic or dielectric-based metasurfaces, and flat-optics-based designer surfaces that can form information processing networks to execute a desired computational task between an input and output aperture.

4.2 Results

Theoretical Analysis of the Information Processing Capacity of Diffractive Surfaces

Let the $\boldsymbol{x}$ and $\boldsymbol{y}$ vectors represent the sampled optical fields (including the phase and amplitude information) at the input and output apertures, respectively. We assume that the sizes of $\boldsymbol{x}$ and $\boldsymbol{y}$ are $N_i \times 1$ and $N_o \times 1$, defined by the input and output fields-of-view, respectively (see Fig. 4.1a); these two quantities, N_i and N_o, are simply proportional to the space-bandwidth

product of the input field and the output field at the input and output apertures of the diffractive network, respectively. Outside the input field-of-view (FOV) defined by N_i, the rest of the points within the input plane do not transmit light or any information to the diffractive network, i.e., they are assumed to be blocked by, for example, an aperture. In a diffractive optical network composed of transmissive and/or reflective surfaces that rely on linear optical materials, these vectors are related to each other by $\boldsymbol{Ax} = \boldsymbol{y}$, where $\boldsymbol{A}$ represents the combined effects of the free-space wave propagation and the transmission through (or reflection off of) the diffractive surfaces, where the size of $\boldsymbol{A}$ is $N_o \times N_i$. The matrix $\boldsymbol{A}$ can be considered the mathematical operator that represents the all-optical processing of the information carried by the input complex field (within the input field-of-view/aperture), delivering the processing results to the desired output field-of-view.

Here, we prove that an optical network having a larger number of diffractive surfaces or trainable neurons can generate a richer set for the transformation matrix $\boldsymbol{A}$ up to a certain limit within the set of all complex-valued matrices with size $N_o \times N_i$. Therefore, this section analytically investigates the all-optical information processing capacity of diffractive networks composed of diffractive surfaces. The input field is assumed to be monochromatic, spatially and temporally coherent with an arbitrary polarization state, and the diffractive surfaces are assumed to be linear, without any coupling to other states of polarization, which is ignored.

Let $\boldsymbol{H_d}$ be an $N \times N$ matrix, which represents the Rayleigh-Sommerfeld diffraction between two fields specified over parallel planes that are axially separated by a distance d. Since $\boldsymbol{H_d}$ is created from the free-space propagation convolution kernel, it is a Toeplitz matrix. Throughout the paper, without loss of generality, we assume that $N_i = N_o = N_{FOV}$, $N \geq N_{FOV}$ and that the

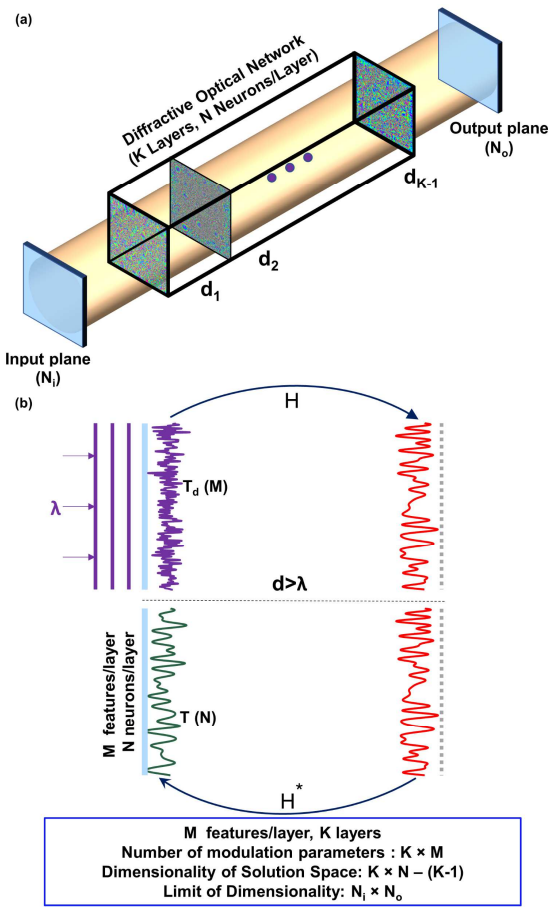

Fig. 4.1 Schematic of a multi-surface diffractive network. a Schematic of a diffractive optical network that connects an input field-of-view (aperture) comprised of N_i points to a desired region-of-interest at the output plane/aperture covering N_o points, through K diffractive surfaces with N neurons per surface, sampled at a period of $\lambda/2n$, where λ and n represent the illumination wavelength and the refractive index of the medium between the surfaces, respectively. Without loss of generality, $n = 1$ has been assumed in this manuscript. b The communication between two successive diffractive surfaces occurs through propagating waves when the axial separation (d) between these layers is larger than λ. Even if the diffractive surface has deeply sub-wavelength structures as in the case of e.g., metasurfaces, with a much smaller sampling period compared to $\lambda/2$ and many more degrees of freedom (M) compared to N, the information processing capability of a diffractive surface within a network is limited to propagating modes since $d > \lambda$; this limits the effective number of neurons per layer to N, even for a surface with $M >> N$. H and H^* refer to the forward and backward wave propagation, respectively.

diffractive surfaces are separated by free space, i.e., the refractive index surrounding the diffractive layers is taken as $n = 1$. We also assume that the optical fields include only the propagating modes, i.e., travelling waves; stated differently, the evanescent modes along the propagation direction are not included in our model since $d \geq \lambda$ (Fig. 4.1b). With this assumption, we choose the sampling period of the discretized complex fields to be $\lambda/2$, where λ is the wavelength of the monochromatic input field. Accordingly, the eigenvalues of $\boldsymbol{H_d}$ are in the form $e^{jk_z d}$ for $0 \leq k_z \leq k_o$, where k_o is the wavenumber of the optical field[133].

Furthermore, let $\boldsymbol{T_k}$ be an $N_{Lk} \times N_{Lk}$ matrix, which represents the k^{th} diffractive surface/layer in the network model, where N_{Lk} is the number of neurons in the corresponding diffractive surface; for a diffractive network composed of K surfaces, without loss of generality we assume $\min(N_{L1}, N_{L2}, \ldots, N_{LK}) \geq N_{FOV}$. Based on these definitions, the elements of $\boldsymbol{T_k}$ are nonzero *only* along its main diagonal entries. These diagonal entries represent the complex-valued transmittance (or reflectance) values (i.e., the optical neurons) of the associated diffractive surface, with a sampling period of $\lambda/2$. Furthermore, each diffractive surface defined by a given transmittance matrix is assumed to be surrounded by a blocking layer within the same plane to avoid any optical communication between the layers without passing through an intermediate diffractive surface. This formalism embraces any form of diffractive surface, including, e.g., plasmonic or dielectric-based metasurfaces. Even if the diffractive surface has deeply sub-wavelength structures, with a much smaller sampling period compared to $\lambda/2$ and many more degrees of freedom (M) compared to N_{Lk}, the information processing capability of a diffractive surface within a network is limited to propagating modes since $d \geq \lambda$, which restricts the effective number of neurons per layer to N_{Lk} (Fig. 4.1b). In other words, since we assume that only propagating modes can reach the subsequent diffractive surfaces within the optical

diffractive network, the sampling period (and hence, the neuron size) of $\lambda/2$ is sufficient to represent these propagating modes in air[134]. According to Shannon's sampling theorem, since the spatial frequency band of the propagating modes in air is restricted to the $(-1/\lambda, 1/\lambda)$ interval, a neuron size that is smaller than $\lambda/2$ leads to oversampling and over-utilization of the optical neurons of a given diffractive surface. On the other hand, if one aims to control and engineer the evanescent modes, then a denser sampling period on each diffractive surface is needed, which might be useful to build diffractive networks that have $d \ll \lambda$. In this near-field diffractive network, the enormously rich degrees of freedom enabled by various metasurface designs with $M \gg N_{Lk}$ can be utilized to provide full and independent control of the phase and amplitude coefficients of each individual neuron of a diffractive surface.

The underlying physical process of how the light is modulated by an optical neuron may vary in different diffractive surface designs. In a dielectric-material-based transmissive design, for example, phase modulation can be achieved by slowing down the light inside the material, where the thickness of an optical neuron determines the amount of phase shift that the light beam undergoes. Alternatively, liquid-crystal (LC)-based spatial light modulators (SLMs) or flat-optics-based metasurfaces can also be employed as part of a diffractive network to generate the desired phase and/or amplitude modulation on the transmitted or reflected light[115,135].

Starting from Section **Error! Reference source not found.**, we investigate the physical properties of $\boldsymbol{A}$, generated by different numbers of diffractive surfaces and trainable neurons. In this analysis, without loss of generality, each diffractive surface is assumed to be transmissive, following the schematics shown in Fig. 4.1a, and its extension to reflective surfaces is straightforward and does not change our conclusions. Finally, multiple (back and forth)

reflections within a diffractive network composed of different layers are ignored in our analysis, as these are much weaker processes compared to the forward propagating modes.

Analysis of a single diffractive surface

The input-output relationship for a single diffractive surface that is placed between an input and an output FOV (Fig. 4.1a) can be written as:

$$\boldsymbol{y} = \boldsymbol{H}'_{\boldsymbol{d_2}}\boldsymbol{T_1}\boldsymbol{H}'_{\boldsymbol{d_1}}\boldsymbol{x} = \boldsymbol{A_1}\boldsymbol{x} \tag{4.10}$$

where $d_1 \geq \lambda$ and $d_2 \geq \lambda$ represent the axial distance between the input plane and the diffractive surface, and the axial distance between the diffractive surface and the output plane, respectively. Here we also assume that $d_1 \neq d_2$; later we discuss the special case of $d_1 = d_2$. Since there is only one diffractive surface in the network, we denote the transmittance matrix as $\boldsymbol{T_1}$, the size of which is $N_{L1} \times N_{L1}$, where $L1$ represents the diffractive surface. Here, $\boldsymbol{H}'_{\boldsymbol{d_1}}$ is an $N_{L1} \times N_{FOV}$ matrix that is generated from the $N_{L1} \times N_{L1}$ propagation matrix $\boldsymbol{H_{d_1}}$ by deleting the appropriately chosen $N_{L1} - N_{FOV}$ -many columns. The positions of the deleted columns correspond to the zero transmission values at the input plane that lie outside the input field-of-view or aperture defined by $N_i = N_{FOV}$ (Fig. 4.1a), i.e., not included in $\boldsymbol{x}$. Similarly, $\boldsymbol{H}'_{\boldsymbol{d_2}}$ is an $N_{FOV} \times N_{L1}$ matrix that is generated from the $N_{L1} \times N_{L1}$ propagation matrix $\boldsymbol{H_{d_2}}$ by deleting the appropriately chosen $N_{L1} - N_{FOV}$ -many rows, which correspond to the locations outside the output FOV or aperture defined by $N_o = N_{FOV}$ in Fig. 4.1a; this means that the output field is calculated only within the desired output aperture. As a result, $\boldsymbol{H}'_{\boldsymbol{d_1}}$ and $\boldsymbol{H}'_{\boldsymbol{d_2}}$ have a rank of N_{FOV}.

To investigate the information processing capacity of $\boldsymbol{A_1}$ based on a single diffractive surface, we vectorize this matrix in the column order and denote it as $vec(\boldsymbol{A_1}) = \boldsymbol{a_1}$ [136]. Next, we show that the set of possible $\boldsymbol{a_1}$ vectors forms a $\min(N_{L1}, N_{FOV}^2)$-dimensional subset of an N_{FOV}^2-dimensional complex-valued vector field. The vector, $\boldsymbol{a_1}$, can be written as:

$$\begin{aligned} vec(\boldsymbol{A_1}) = \boldsymbol{a_1} &= vec\left(\boldsymbol{H'_{d_2}} \boldsymbol{T_1} \boldsymbol{H'_{d_1}}\right) \\ &= \left(\boldsymbol{H'^{T}_{d_1}} \otimes \boldsymbol{H'_{d_2}}\right) vec(\boldsymbol{T_1}) \\ &= \left(\boldsymbol{H'^{T}_{d_1}} \otimes \boldsymbol{H'_{d_2}}\right) \boldsymbol{t_1} \end{aligned} \tag{4.2}$$

where the superscript T and $\otimes$ denote the transpose operation and Kronecker product, respectively[136]. Here, the size of $\boldsymbol{H'^{T}_{d_1}} \otimes \boldsymbol{H'_{d_2}}$ is $N_{FOV}^2 \times N_{L1}^2$, and it is a full-rank matrix with rank N_{FOV}^2. In Equation 4.2, $vec(\boldsymbol{T_1}) = \boldsymbol{t_1}$ has at most N_{L1} controllable/adjustable complex-valued entries, which physically represent the neurons of the diffractive surface, and the rest of its entries are all zero. These transmission coefficients lead to a linear combination of N_{L1}-many vectors of $\boldsymbol{H'^{T}_{d_1}} \otimes \boldsymbol{H'_{d_2}}$, where $d_1 \neq d_2 \neq 0$. If $N_{L1} \leq N_{FOV}^2$, these vectors subject to the linear combination are linearly independent (see Figure 4.2). Hence, the set of resulting $\boldsymbol{a_1}$ vectors generated by Equation 4.2 forms an N_{L1}-dimensional subspace of the N_{FOV}^2-dimensional complex-valued vector space. On the other hand, if $N_{L1} > N_{FOV}^2$, then the vectors in the linear combination start to become dependent on each other. In this case of $N_{L1} > N_{FOV}^2$, the dimensionality of the set of possible vector fields is limited to N_{FOV}^2 (also see Figure 4.2).

This analysis demonstrates that the set of complex field transformation vectors that can be generated by a single diffractive surface that connects a given input and output FOV constitutes a $\min(N_{L1}, N_{FOV}^2)$-dimensional subspace of an N_{FOV}^2-dimensional complex-valued vector space. These results are based on our earlier assumption that $d_1 \geq \lambda$, $d_2 \geq \lambda$ and $d_1 \neq d_2$. For the special case of $d_1 = d_2 \geq \lambda$, the upper limit of the dimensionality of the solution space that can be

(a) K=1, $N_i = N_o = N_{FOV} = 4 \times 4$

$D = min(N_{L1},\ N_i \times N_o)$	$N_{L1} = N_{y1} \times N_{x1}$						
	20 x 20	16 x 16	12 x 12	12 x 11	11 x 12	17 x 8	8 x 17
$d_1 = \lambda,\ d_2 = 4\lambda$	256	256	144	132	132	136	136
$d_1 = 4\lambda,\ d_2 = \lambda$	256	256	144	132	132	136	136
$d_1 = 4\lambda,\ d_2 = 64\lambda$	256	256	144	132	132	136	136
$d_1 = 64\lambda,\ d_2 = 4\lambda$	256	256	144	132	132	136	136
$d_1 = 1024\lambda,\ d_2 = 16\lambda$	256	256	144	132	132	136	136
$d_1 = 16\lambda,\ d_2 = 1024\lambda$	256	256	144	132	132	136	136
$d_1 = 64\lambda,\ d_2 = 1\lambda$	256	256	144	132	132	136	136
$d_1 = 1\lambda,\ d_2 = 64\lambda$	256	256	144	132	132	136	136

(b) K=1, $N_i = N_o = N_{FOV} = 8 \times 8$

$D = min(N_{L1},\ N_i \times N_o)$	D
$d_1 = 1\lambda,\ d_2 = 4\lambda,\ N_{L1} = 64 \times 64$	4096
$d_1 = 4\lambda,\ d_2 = 64\lambda,\ N_{L1} = 64 \times 64$	4096
$d_1 = 1024\lambda,\ d_2 = 16\lambda,\ N_{L1} = 64 \times 64$	4096
$d_1 = 64\lambda,\ d_2 = 1\lambda,\ N_{L1} = 64 \times 64$	4096
$d_1 = 1\lambda,\ d_2 = 4\lambda,\ N_{L1} = 46 \times 46$	2116
$d_1 = 4\lambda,\ d_2 = 64\lambda,\ N_{L1} = 46 \times 46$	2116
$d_1 = 1024\lambda,\ d_2 = 16\lambda,\ N_{L1} = 46 \times 46$	2116
$d_1 = 64\lambda,\ d_2 = 1\lambda,\ N_{L1} = 46 \times 46$	2116
$d_1 = 4\lambda,\ d_2 = 64\lambda,\ N_{L1} = 40 \times 52$	2080
$d_1 = 4\lambda,\ d_2 = 64\lambda,\ N_{L1} = 52 \times 40$	2080
$d_1 = 4\lambda,\ d_2 = 64\lambda,\ N_{L1} = 20 \times 104$	2080
$d_1 = 4\lambda,\ d_2 = 64\lambda,\ N_{L1} = 104 \times 20$	2080
$d_1 = 1\lambda,\ d_2 = 4\lambda,\ N_{L1} = 32 \times 128$	4096
$d_1 = 1\lambda,\ d_2 = 4\lambda,\ N_{L1} = 128 \times 32$	4096

Fig. 4.2: Computation of the dimensionality (D) of the all-optical solution space for K=1 diffractive surface under various network configurations. The rank values are obtained using the symbolic toolbox of MATLAB from H' matrix. The calculated rank values in each table obey the rule $D = min(N_{L1}, N_{FOV}^2)$. $D = N_{L1}$ results indicate that all the columns of H' are linearly independent, and therefore any subset of its columns are also linearly independent. Therefore, the dimensionality of the solution space for $d_1 \neq d_2$ is a linear function of N_{L1} when $N_{L1} \leq N_{FOV}^2$, and N_{FOV}^2 defines the upper limit of D (see Figure 4.5). We also show that the upper limit for the dimensionality of the all-optical solution space reduces to $N_{FOV}(N_{FOV}+1)/2$ when $d_1 = d_2$ for a single diffractive layer, $K = 1$.

generated by a single diffractive surface (K=1) is reduced from N_{FOV}^2 to $(N_{FOV}^2 + N_{FOV})/2$ due to the combinatorial symmetries that exist in the optical path for $d_1 = d_2$.

Analysis of an optical network formed by two diffractive surfaces

Here, we consider an optical network with two different (trainable) diffractive surfaces (K=2), where the input-output relation can be written as:

$$\boldsymbol{y} = \boldsymbol{H'_{d_3} T_2 H_{d_2} T_1 H'_{d_1} x} = \boldsymbol{A_2 x} \tag{4.11}$$

$N_x = \max(N_{L1}, N_{L2})$ determines the sizes of the matrices in Equation 4.3, where N_{L1} and N_{L2} represent the number of neurons in the first and second diffractive surfaces, respectively; d_1, d_2 and d_3 represent the axial distances between the diffractive surfaces (see Fig. 4.1a). Accordingly, the sizes of $\boldsymbol{H'_{d_1}}$, $\boldsymbol{H_{d_2}}$ and $\boldsymbol{H'_{d_3}}$ become $N_x \times N_{FOV}$, $N_x \times N_x$ and $N_{FOV} \times N_x$, respectively. Since we have already assumed that $\min(N_{L1}, N_{L2}) \geq N_{FOV}$, $\boldsymbol{H'_{d_1}}$ and $\boldsymbol{H'_{d_3}}$ can be generated from the corresponding $N_x \times N_x$ propagation matrices by deleting the appropriate columns and rows, as described in Section **Error! Reference source not found.**. Because $\boldsymbol{H_{d_2}}$ has a size of $N_x \times N_x$, there is no need to delete any rows or columns from the associated propagation matrix. Although both $\boldsymbol{T_1}$ and $\boldsymbol{T_2}$ have a size of $N_x \times N_x$, the one corresponding to the diffractive surface that contains the smaller number of neurons has some zero values along its main diagonal indices. The number of these zeros is $N_x - \min(N_{L1}, N_{L2})$.

Similar to the analysis reported in Section **Error! Reference source not found.**, the vectorization of $\boldsymbol{A_2}$ reveals:

$$\begin{aligned} vec(\boldsymbol{A_2}) = \boldsymbol{a_2} &= vec\left(\boldsymbol{H'_{d_3} T_2 H_{d_2} T_1 H'_{d_1}}\right) \\ &= \left(\boldsymbol{H'^{T}_{d_1}} \otimes \boldsymbol{H'_{d_3}}\right) vec\left(\boldsymbol{T_2 H_{d_2} T_1}\right) \\ &= \left(\boldsymbol{H'^{T}_{d_1}} \otimes \boldsymbol{H'_{d_3}}\right)\left(\boldsymbol{T_1^T} \otimes \boldsymbol{T_2}\right) vec\left(\boldsymbol{H_{d_2}}\right) \end{aligned} \tag{4.4}$$

$$= \left(\boldsymbol{H}_{\boldsymbol{d_1}}^{\prime T} \otimes \boldsymbol{H}_{\boldsymbol{d_3}}^{\prime}\right)(\boldsymbol{T_1} \otimes \boldsymbol{T_2}) vec\left(\boldsymbol{H}_{\boldsymbol{d_2}}\right)$$
$$= \left(\boldsymbol{H}_{\boldsymbol{d_1}}^{\prime T} \otimes \boldsymbol{H}_{\boldsymbol{d_3}}^{\prime}\right)(\boldsymbol{T_1} \otimes \boldsymbol{T_2}) \boldsymbol{h}_{\boldsymbol{d_2}}$$
$$= \left(\boldsymbol{H}_{\boldsymbol{d_1}}^{\prime T} \otimes \boldsymbol{H}_{\boldsymbol{d_3}}^{\prime}\right) \widehat{\boldsymbol{H}}_{\boldsymbol{d_2}} diag(\boldsymbol{T_1} \otimes \boldsymbol{T_2})$$
$$= \left(\boldsymbol{H}_{\boldsymbol{d_1}}^{\prime T} \otimes \boldsymbol{H}_{\boldsymbol{d_3}}^{\prime}\right) \widehat{\boldsymbol{H}}_{\boldsymbol{d_2}} \boldsymbol{t_{12}}$$

where $\widehat{\boldsymbol{H}}_{\boldsymbol{d_2}}$ is an $N_x^2 \times N_x^2$ matrix that has nonzero entries *only* along its main diagonal locations. These entries are generated from $vec\left(\boldsymbol{H}_{\boldsymbol{d_2}}\right) = \boldsymbol{h}_{\boldsymbol{d_2}}$ such that $\widehat{\boldsymbol{H}}_{\boldsymbol{d_2}}[i,i] = \boldsymbol{h}_{\boldsymbol{d_2}}[i]$. Since the $diag(\cdot)$ operator forms a vector from the main diagonal entries of its input matrix, the vector $\boldsymbol{t_{12}} = diag(\boldsymbol{T_1} \otimes \boldsymbol{T_2})$ is generated such that $\boldsymbol{t_{12}}[i] = (\boldsymbol{T_1} \otimes \boldsymbol{T_2})[i,i]$. The equality $(\boldsymbol{T_1} \otimes \boldsymbol{T_2})\boldsymbol{h}_{\boldsymbol{d_2}} = \widehat{\boldsymbol{H}}_{\boldsymbol{d_2}} \boldsymbol{t_{12}}$ stems from the fact that the nonzero elements of $\boldsymbol{T_1} \otimes \boldsymbol{T_2}$ are located only along its main diagonal entries.

In Equation 4.4, $\boldsymbol{H}_{\boldsymbol{d_1}}^{\prime T} \otimes \boldsymbol{H}_{\boldsymbol{d_3}}^{\prime}$ has rank N_{FOV}^2. Since all the diagonal elements of $\widehat{\boldsymbol{H}}_{\boldsymbol{d_2}}$ are nonzero, it has rank N_x^2. As a result, $\left(\boldsymbol{H}_{\boldsymbol{d_1}}^{T} \otimes \boldsymbol{H}_{\boldsymbol{d_3}}\right)\widehat{\boldsymbol{H}}_{\boldsymbol{d_2}}$ is a full-rank matrix with rank N_{FOV}^2. Additionally, the nonzero elements of $\boldsymbol{t_{12}}$ take the form $t_{ij} = t_{1,i} t_{2,j}$, where $t_{1,i}$ and $t_{2,j}$ are the trainable/adjustable complex transmittance values of the i^{th} neuron of the 1st diffractive surface and the j^{th} neuron of the 2nd diffractive surface, respectively, for $i \in \{1,2,\dots,N_{L1}\}$ and $j \in \{1,2,\dots,N_{L2}\}$. Then, the set of possible $\boldsymbol{a_2}$ vectors (Equation 4.4) can be written as:

$$\boldsymbol{a_2} = \sum_{i,j} t_{ij} \boldsymbol{h_{ij}} \tag{4.5}$$

where $\boldsymbol{h_{ij}}$ is the corresponding column vector of $\left(\boldsymbol{H}_{\boldsymbol{d_1}}^{\prime T} \otimes \boldsymbol{H}_{\boldsymbol{d_3}}^{\prime}\right)\widehat{\boldsymbol{H}}_{\boldsymbol{d_2}}$.

Equation 4.5 is in the form of a complex-valued linear combination of $N_{L1}N_{L2}$-many complex-valued vectors, $\boldsymbol{h_{ij}}$. Since we assume $\min(N_{L1}, N_{L2}) \geq N_{FOV}$, these vectors necessarily form a linearly dependent set of vectors and this restricts the dimensionality of the vector space to N_{FOV}^2. Moreover, due to the coupling of the complex-valued transmittance values of the two

diffractive surfaces ($t_{ij} = t_{1,i}t_{2,j}$) in Equation 4.5, the dimensionality of the resulting set of $\boldsymbol{a_2}$ vectors can even go below N_{FOV}^2, despite $N_{L1}N_{L2} \geq N_{FOV}^2$. In fact, in the Materials and Methods section, we show that the set of $\boldsymbol{a_2}$ vectors can form an $N_{L1}+N_{L2}-1$-dimensional subspace of the N_{FOV}^2-dimensional complex-valued vector space and can be written as:

$$\boldsymbol{a_2} = \sum_{k=1}^{N_{L1}+N_{L2}-1} c_k \boldsymbol{b_k} \tag{4.6}$$

where $\boldsymbol{b_k}$ represents length-N_{FOV}^2 linearly independent vectors and c_k represents complex-valued coefficients, generated through the coupling of the transmittance values of the two independent diffractive surfaces. The relationship between Equations 4.5 and 4.6 is also presented as a pseudo-code in Tables 4.1-4.3 and Figure 4.3.

These analyses reveal that by using a diffractive optical network composed of two different trainable diffractive surfaces (with neurons N_{L1}, N_{L2}), it is possible to generate an all-optical solution that spans an $N_{L1}+N_{L2}-1$ dimensional subspace of an N_{FOV}^2-dimensional complex-valued vector space. As a special case, if we assume $N = N_{L1} = N_{L2} = N_i = N_o = N_{FOV}$, the resulting set of complex-valued linear transformation vectors forms a $2N-1$ dimensional subspace of an N^2-dimensional vector field. Table 4.1 also provides a coefficient and basis vector generation algorithm, independently reaching the same conclusion that this special case forms a $2N-1$ dimensional subspace of an N^2-dimensional vector field. The upper limit of the solution space dimensionality that can be achieved by a two-layered diffractive network is N_{FOV}^2, which is dictated by the input and output fields-of-view between which the diffractive network is positioned.

(a) $K=2$, $N_i = N_o = N_{FOV} = 4 \times 4$

$D = min(N_{L1} + N_{L2} - 1, N_i \times N_o)$	$N_{L1} = N_{y1} \times N_{x1}$, $N_{L2} = N_{y2} \times N_{x2}$						
	$N_{L1} = 16 \times 10$, $N_{L2} = 10 \times 16$	$N_{L1} = 16 \times 8$, $N_{L2} = 8 \times 16$	$N_{L1} = 14 \times 14$, $N_{L2} = 7 \times 7$	$N_{L1} = 7 \times 7$, $N_{L2} = 14 \times 14$	$N_{L1} = 8 \times 8$, $N_{L2} = 10 \times 10$	$N_{L1} = 10 \times 10$, $N_{L2} = 8 \times 8$	$N_{L1} = 11 \times 11$, $N_{L2} = 11 \times 11$
$d_1 = d_2 = d_3 = \lambda$	256	255	244	244	163	163	241
$d_1 = d_2 = d_3 = 4\lambda$	256	255	244	244	163	163	241
$d_1 = \lambda, d_2 = 4\lambda, d_3 = 16\lambda$	256	255	244	244	163	163	241
$d_1 = d_2 = d_3 = 10^3 \lambda$	256	255	244	244	163	163	241
$d_1 = d_2 = d_3 = 4 \times 10^3 \lambda$	256	255	244	244	163	163	241
$d_1 = 10^3 \lambda, d_2 = 4 \times 10^3 \lambda, d_3 = 16 \times 10^3 \lambda$	256	255	244	244	163	163	241

(b) $K=2$, $N_i = N_o = N_{FOV} = 6 \times 6$ and 7×7

$D = min(N_{L1} + N_{L2} - 1, N_i \times N_o)$		D
$N_i = N_o = 6 \times 6$, $N_{L1} = 36 \times 18$, $N_{L2} = 18 \times 36$	$d_1 = d_2 = d_3 = \lambda$	1295
	$d_1 = d_2 = d_3 = 4\lambda$	1295
	$d_1 = 1\lambda, d_2 = 4\lambda, d_3 = 16\lambda$	1295
	$d_1 = d_2 = d_3 = 10^3 \lambda$	1295
	$d_1 = d_2 = d_3 = 4 \times 10^3 \lambda$	1295
	$d_1 = 10^3 \lambda, d_2 = 4 \times 10^3 \lambda, d_3 = 16 \times 10^3 \lambda$	1295
$N_i = N_o = 7 \times 7$, $N_{L1} = 32 \times 48$, $N_{L2} = 48 \times 32$	$d_1 = d_2 = d_3 = \lambda$	2401
	$d_1 = d_2 = d_3 = 4\lambda$	2401
	$d_1 = 1\lambda, d_2 = 4\lambda, d_3 = 16\lambda$	2401
	$d_1 = d_2 = d_3 = 10^3 \lambda$	2401
	$d_1 = d_2 = d_3 = 4 \times 10^3 \lambda$	2401
	$d_1 = 10^3 \lambda, d_2 = 4 \times 10^3 \lambda, d_3 = 16 \times 10^3 \lambda$	2401

(c) $K=2$, $N_i = N_o = N_{FOV} = 8 \times 8$

$D = min(N_{L1} + N_{L2} - 1, N_i \times N_o)$	$N_{L1} = N_{y1} \times N_{x1}$, $N_{L2} = N_{y2} \times N_{x2}$	
	$N_{L1} = 24 \times 48$, $N_{L2} = 48 \times 24$	$N_{L1} = 24 \times 24$, $N_{L2} = 12 \times 12$
$d_1 = d_2 = d_3 = \lambda$	2303	719
$d_1 = d_2 = d_3 = 4\lambda$	2303	719
$d_1 = 1\lambda, d_2 = 4\lambda, d_3 = 16\lambda$	2303	719
$d_1 = d_2 = d_3 = 10^3 \lambda$	2303	719
$d_1 = d_2 = d_3 = 4 \times 10^3 \lambda$	2303	719
$d_1 = 10^3 \lambda, d_2 = 4 \times 10^3 \lambda, d_3 = 16 \times 10^3 \lambda$	2303	719

Fig. 4.3: Computation of the dimensionality (D) of the all-optical solution space for K=2 diffractive surfaces under various network configurations. The rank values are obtained using the symbolic toolbox of MATLAB from H' matrix. Each result in the presented tables is confirmed through three independent runs of the same algorithm with different random initializations, random selection of the neurons and random generation of complex-valued transmission coefficients. All the presented rank results numerically confirm $D = min(N_{L1} + N_{L2} - 1, N_{FOV}^2)$.

In summary, these analyses showed that the dimensionality of the all-optical solution space covered by two trainable diffractive surfaces (K=2) positioned between a given set of input-output FOV is given by $\min(N_{FOV}^2, \mathrm{N}_{L1}+\mathrm{N}_{L2}-1)$. Different from K=1 architecture, which revealed a restricted solution space when $d_1 = d_2$, diffractive optical networks with K=2 do not exhibit a similar restriction related to the axial distances d_1, d_2 and d_3 (see Fig. 4.3).

Analysis of an optical network formed by three or more diffractive surfaces

Next, we consider an optical network formed by more than two diffractive surfaces, with neurons of $(N_{L1}, N_{L2}, \cdots, N_{LK})$ for each layer, where K is the number of diffractive surfaces and N_{Lk} represents the number of neurons in the k^{th} layer. In the previous section, we showed that a two-layered network with (N_{L1}, N_{L2}) neurons has the same solution space dimensionality as that of a single-layered, larger diffractive network having $\mathrm{N}_{L1}+\mathrm{N}_{L2}-1$ individual neurons. If we assume that a third diffractive surface (N_{L3}) is added to this single-layer network with $\mathrm{N}_{L1}+\mathrm{N}_{L2}-1$ neurons, this becomes equivalent to a two-layered network with $(N_{L1}+N_{L2}-1, N_{L3})$ neurons. The dimensionality of the all-optical solution space covered by this diffractive network positioned between a set of input-output fields-of-view is given by $\min(N_{FOV}^2, N_{L1}+N_{L2}+N_{L3}-2)$; also see Fig. 4.4. For the special case of $N_{L1} = N_{L2} = N_{L3} = N_i = N_o = N$.

The above arguments can be extended to a network that has K diffractive surfaces. That is, for a multi-surface diffractive network with a neuron distribution of $(N_{L1}, N_{L2}, \cdots, N_{LK})$, the dimensionality of the solution space (see Fig. 4.5) created by this diffractive network is given by:

$$\min\left(N_{FOV}^2, \left[\sum_{k=1}^{K} N_{Lk}\right] - (K-1)\right) \tag{4.7}$$

which forms a subspace of an N_{FOV}^2-dimensional vector space that covers all the complex-valued linear transformations between the input and output fields-of-view.

The upper bound on the dimensionality of the solution space, i.e., the N_{FOV}^2 term in Equation 4.7, is heuristically imposed by the number of possible ray interactions between the input and output fields-of-view. That is, if we consider the diffractive optical network as a black box (Fig. 4.1a), its operation can be intuitively understood as controlling the phase and/or amplitude of the light rays that are collected from the input, to be guided to the output, following a lateral grid of $\lambda/2$ at the input/output fields-of-view, determined by the diffraction limit of light. The second term in Equation 4.7, on the other hand, reflects the total space-bandwidth product of K successive diffractive surfaces, one following another. To intuitively understand the $(K-1)$ subtraction term in Equation 4.7, one can hypothetically consider the simple case of $N_{Lk} = N_{FOV} = 1$ for all K diffractive layers; in this case, $[\sum_{k=1}^{K} N_{Lk}] - (K-1) = 1$, which simply indicates that K successive diffractive surfaces (each with $N_{Lk} = 1$) are equivalent, as physically expected, to a single controllable diffractive surface with N_L=1.

Without loss of generality, if we assume $N = N_k$ for all the diffractive surfaces, then the dimensionality of the linear transformation solution space created by this diffractive network will be $KN - (K-1)$, provided that $KN - (K-1) \leq N_{FOV}^2$. This means that for a fixed design choice of N neurons per diffractive surface (determined by, e.g., the limitations of the fabrication methods or other practical considerations), adding new diffractive surfaces to the same

diffractive network linearly increases the dimensionality of the solution space that can be all-optically processed by the diffractive network between the input/output fields-of-view. As we further increase K such that $KN-(K-1) \geq N_{FOV}^2$, the diffractive network reaches its linear transformation capacity, and adding more layers or more neurons to the network does not further contribute to its processing power for the desired input-output fields-of-view (see Fig. 4.5). However, these deeper diffractive networks that have larger numbers of diffractive surfaces (i.e., $KN-(K-1) \geq N_{FOV}^2$) can cover a solution space with a dimensionality of $KN-(K-1)$ over larger input and output fields-of-view. Stated differently, for any given choice of N neurons per diffractive surface, deeper diffractive networks that are composed of multiple surfaces can cover a $KN-(K-1)$-dimensional subspace of all the complex-valued linear transformations between a larger input field-of-view ($N'_i > N_i$) and/or a larger output field-of view ($N'_o > N_o$), as long as $KN-(K-1) \leq N'_i N'_o$. The conclusions of this analysis are also summarized in Fig. 4.5.

In addition to increasing K (the number of diffractive surfaces within an optical network), an alternative strategy to increase the all-optical processing capabilities of a diffractive network is to increase N, the number of neurons per diffractive surface/layer. However, as we numerically demonstrate in the next section, this strategy is not as effective as increasing the number of

$K=3, \quad N_i = N_o = N_{FOV} = 4 \times 4$

$D = min(N_{L1} + N_{L2} + N_{L3} - 2, N_i \times N_o)$	$N_{L1} = N_{y1} \times N_{x1}$ $N_{L2} = N_{y2} \times N_{x2}$ $N_{L3} = N_{y3} \times N_{x3}$	
	$N_{L1} = 6 \times 12$ $N_{L2} = 18 \times 6$ $N_{L3} = 18 \times 12$	$N_{L1} = 16 \times 8$ $N_{L2} = 8 \times 6$ $N_{L3} = 8 \times 10$
$d_1 = d_2 = d_3 = d_4 = \lambda$	256	254
$d_1 = d_2 = d_3 = d_4 = 4\lambda$	256	254
$d_1 = \lambda,\ d_2 = 4\lambda,\ d_3 = 8\lambda,\ d_4 = 16\lambda$	256	254
$d_1 = d_2 = d_3 = d_4 = 10^3 \lambda$	256	254
$d_1 = d_2 = d_3 = d_4 = 4 \times 10^3 \lambda$	256	254
$d_1 = 10^3 \lambda,\ d_2 = 4 \times 10^3 \lambda,\ d_3 = 8 \times 10^3 \lambda,\ d_4 = 16 \times 10^3 \lambda$	256	254

Fig. 4.4: Computation of the dimensionality (D) of the all-optical solution space for K=3 diffractive surfaces under various network configurations. The rank values are obtained using the symbolic toolbox of MATLAB from H' matrix. The presented results indicate that it is possible to obtain the maximum dimensionality of the solution space, numerically confirming that $D = min(N_{L1} + N_{L2} + N_{L3} - 2, N_{FOV}^2)$.

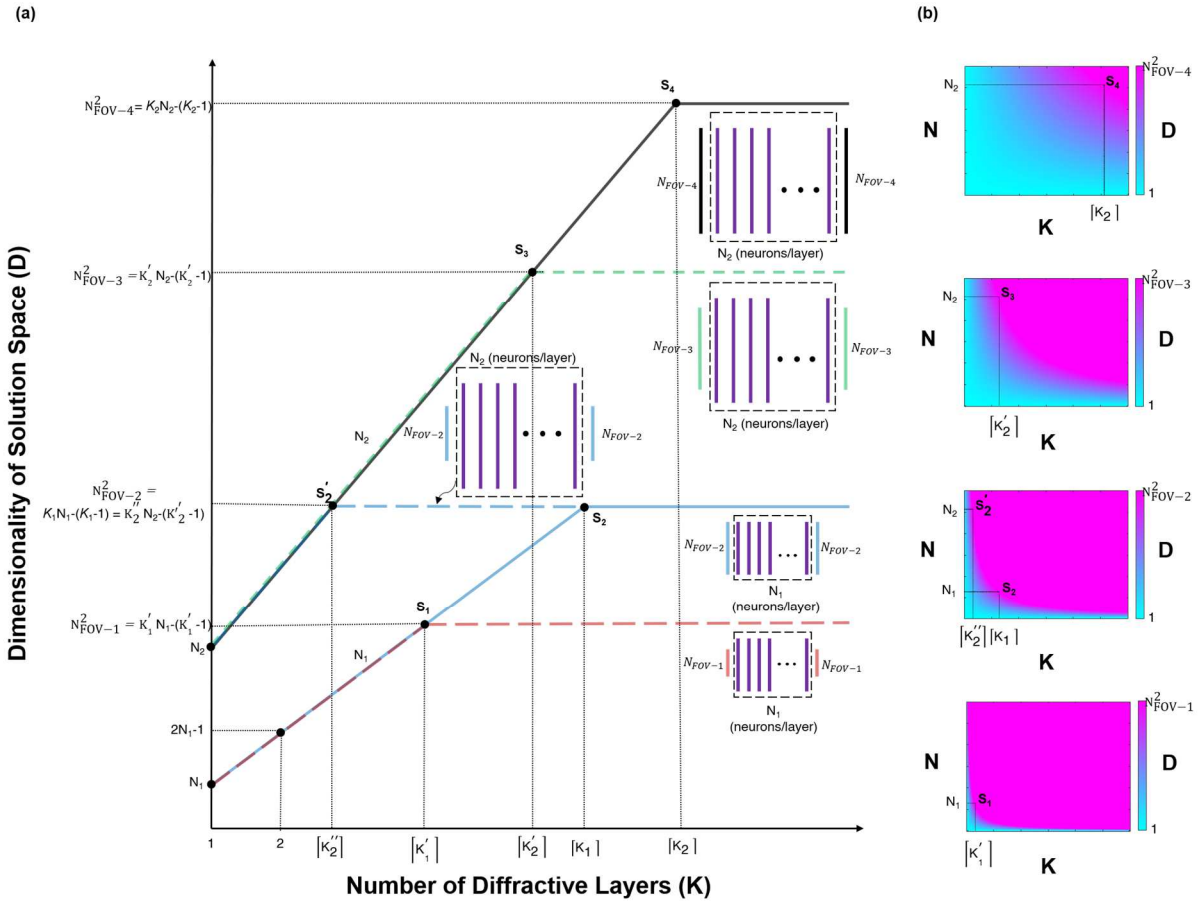

Fig. 4.5: Dimensionality (D) of the all-optical solution space covered by multi-layer diffractive networks. a The behavior of the dimensionality of the all-optical solution space with increasing number of layers for two different diffractive surface designs, with $N = N_1$ and $N = N_2$ neurons per surface, where $N_2 > N_1$. The smallest number of diffractive surfaces, $\lceil K_s \rceil$, satisfying the condition, $K_S N - (K_S - 1) \geq N_i \times N_o$, determines the ideal depth of the network for a given N, N_i and N_O. For the sake of simplicity, here we assumed $N_i = N_o = N_{FOV-i}$, where 4 different input/output fields-of-view are illustrated in the plot, i.e., $N_{FOV-4} > N_{FOV-3} > N_{FOV-2} > N_{FOV-1}$. $\lceil K_s \rceil$ refers to the ceiling function, defining the number of diffractive surfaces within an optical network design. b The distribution of the dimensionality of the all-optical solution space as a function of N and K for 4 different field-of-views, N_{FOV-i}, and the corresponding turning points, S_i, which are shown in (a).

diffractive surfaces since deep-learning-based design tools are relatively inefficient in utilizing all the degrees of freedom provided by a diffractive surface with $N >> N_o, N_i$. This is partially related to the fact that high-numerical-aperture optical systems are generally more difficult to optimize and design. Moreover, if we consider a single-layer diffractive network design with a large N_{max} (which defines the *maximum* surface area that can be fabricated and engineered with the desired transmission coefficients), even for this N_{max} design, the addition of new diffractive surfaces with N_{max} at each surface linearly increases the dimensionality of the solution space created by the diffractive network, covering linear transformations over larger input and output fields-of-view, as discussed earlier. These reflect some of the important depth advantages of diffractive optical networks that are formed by multiple diffractive surfaces. The next section further expands on this using a numerical analysis of diffractive optical networks that are designed for image classification.

Computation of the dimensionality (D) of the all-optical solution space for K = 1, 2 and 3

In order to calculate D for various diffractive network configurations, we used the symbolic toolbox of MATLAB to compute the rank of diffraction related matrices using their symbolic representation.

1-Layer Case (K = 1):

To compute D for $K = 1$, we first generate $\boldsymbol{H}_{\boldsymbol{d_1}}^{\prime T} \otimes \boldsymbol{H}_{\boldsymbol{d_2}}^{\prime}$ of Equation 4.2. Note that for $K = 1$ only N_{L1} −many columns of $\boldsymbol{H}_{\boldsymbol{d_1}}^{\prime T} \otimes \boldsymbol{H}_{\boldsymbol{d_2}}^{\prime}$ are included in the computation of $vec(\boldsymbol{A_1})$. Therefore, we consider only those vectors in our computation. We define $\boldsymbol{H}'$ as the matrix which is subject to the rank computation:

$$\boldsymbol{H}'[:,m] = (\boldsymbol{H}'^{T}_{\boldsymbol{d_1}} \otimes \boldsymbol{H}'_{\boldsymbol{d_2}})[:, m(N_{L1}+1)] \tag{4.8}$$

for $m \in \{0,1,\dots,N_{L1}-1\}$. Here, $[:, m(N_{L1}+1)]$ indicates the column associated with the $(m+1)^{th}$ neuron in the vectorized form. Hence, in 2D discrete space, m corresponds to a certain neuron position and discrete index, $[q_{L1}, p_{L1}]$.

$\boldsymbol{H}'[l,m]$ takes its values through the multiplication of the appropriate free space impulse response functions from the associated input pixel (within N_i) to the $(m+1)^{th}$ neuron and from the $(m+1)^{th}$ neuron to the associated output pixel (within N_o). Thus, a given l corresponds to a certain position at the input plane, $[q_i, p_i]$, paired with a certain position at the output plane, $[q_o, p_o]$. As a result, $\boldsymbol{H}'[l,m]$ can be written as:

$$\boldsymbol{H}'[l,m] = h_{d1}(q_i - q_{L1}, p_i - p_{L1}) \cdot h_{d2}(q_o - q_{L1}, p_o - p_{L1}), \tag{4.9}$$

where $d_1 \neq d_2 \neq 0$ and $h_d(x,y)$ is the impulse response of free space propagation, which can be written as:

$$h_d(x,y) = -\frac{e^{j\frac{2\pi}{\lambda}r}}{2\pi}\left(j\frac{2\pi}{\lambda} - \frac{1}{r}\right)\frac{d}{r^2} \quad , \tag{4.10}$$

where $r = \sqrt{x^2 + y^2 + d^2}$.

In MATLAB, we used various symbolic conversion schemes to confirm that each method ends up with the same rank. For a given $N_i = N_o = N_{FOV}$, N_{L1}, d_1 and d_2 configuration, in the first four methods, we generated $\boldsymbol{H}'$ numerically in the double precision. Then we converted it to the corresponding symbolic matrix representation using either one of these commands:

*>>sym(**H**′,′r′) (Method 1.a)*

*>>sym(**H**′,′d′) (Method 1.b)*

*>>sym(**H**′,′e′) (Method 1.c)*

*>>sym(**H**′,′f′) (Method 1.d)*

In the second set of symbolic conversion schemes, in order to further increase the precision in our computation, we generated π symbolically at the beginning as:

>>sym(pi,′r′) (Method 2.a)

>>sym(pi,′d′) (Method 2.b)

>>sym(pi,′e′) (Method 2.c)

>>sym(pi,′f′) (Method 2.d)

Then we generated $\boldsymbol{H}'$ of Equation 4.8 using the symbolic π, which ended up with a symbolic $\boldsymbol{H}'$ matrix. Note that, although the second set of methods has a better accuracy in symbolic representation, they require more computation memory and time in generating the rank result. So, in our rank computations, we used *Method 1.a* as the common method for all the diffractive network configurations reported in Figs. 4.2-4.4. Besides *Method 1.a,* we also used at least one of the remaining seven methods in each diffractive network configuration to confirm that the resulting rank values agree with each other.

Figure 4.2 summarizes the resulting rank calculations for various different $K = 1$ diffractive network configurations, all of which confirm $D = min(\, N_{L1}\, , N_{FOV}^2\,)$. $D = N_{L1}$ results reported in

Figure 4.2 indicate that all the columns of $\boldsymbol{H'}$ are linearly independent, and therefore any subset of its columns are also linearly independent. This shows that the dimensionality of the solution space for $d_1 \neq d_2$ is a linear function of N_{L1} when $N_{L1} \leq N_{FOV}^2$, and N_{FOV}^2 defines the upper limit of D (also see Fig. 4.5). We also show that the upper limit for the dimensionality of the all-optical solution space reduces to $N_{FOV}(N_{FOV}+1)/2$ when $d_1 = d_2$ for a single diffractive layer, $K = 1$.

2-Layer Case ($K = 2$):

For $K = 2$, we deal with the matrix $\left(\boldsymbol{H}_{\boldsymbol{d_1}}^{\boldsymbol{'T}} \otimes \boldsymbol{H}_{\boldsymbol{d_3}}^{\boldsymbol{'}}\right)\widehat{\boldsymbol{H}}_{\boldsymbol{d_2}}$ of Equation 4.4. We first generated a matrix $\boldsymbol{H'}$ from $\left(\boldsymbol{H}_{\boldsymbol{d_1}}^{\boldsymbol{'T}} \otimes \boldsymbol{H}_{\boldsymbol{d_3}}^{\boldsymbol{'}}\right)\widehat{\boldsymbol{H}}_{\boldsymbol{d_2}}$ such that the columns of $\left(\boldsymbol{H}_{\boldsymbol{d_1}}^{\boldsymbol{'T}} \otimes \boldsymbol{H}_{\boldsymbol{d_3}}^{\boldsymbol{'}}\right)\widehat{\boldsymbol{H}}_{\boldsymbol{d_2}}$ that correspond to the zero entries of $\boldsymbol{t_{12}}$ are discarded. First, we converted $\boldsymbol{H'}$ into a symbolic matrix and then applied the algorithm presented in Table 4.1 on the columns of $\boldsymbol{H'}$. Here the m^{th} column of $\boldsymbol{H'}$ is the vector that multiplies the coefficient $t_{1,i}t_{2,j}$ of $\boldsymbol{t_{12}}$ of Equation 4.4 for a certain (i,j) pair, i.e., there is a one-to-one relationship between a given m and the associated (i,j) pair.

Therefore, a given m indicates a certain neuron position in the first diffractive layer, $[q_{L1}, p_{L1}]$, paired with a certain neuron position in the second diffractive layer, $[q_{L2}, p_{L2}]$. Similar to the $K = 1$ case, the l^{th} row of $\boldsymbol{H'}$ corresponds to a certain set of input and output pixels as part of N_i and N_o, respectively, and $\boldsymbol{H'}[l,m]$ can be written as:

$$\boldsymbol{H'}[l,m] = h_{d1}(q_i - q_{L1}, p_i - p_{L1}) \cdot h_{d3}(q_o - q_{L2}, p_o - p_{L2}) \cdot h_{d2}(q_{L1} - q_{L2}, p_{L1} - p_{L2}) \tag{4.11}$$

After generating $\boldsymbol{H'}$ based on Equation 4.11, we converted it into the symbolic matrix representation as described earlier for the 1-layer case, $K = 1$. Then, we applied the algorithm

presented in Table 4.1 to generate the basis vectors and their coefficients. Note that, for each diffractive network configuration that we selected, we independently ran the same algorithm three times with different random initializations, random selection of the neurons and random generation of complex-valued transmission coefficients. In all of the rank results that are reported in Fig. 4.3, these repeated simulations agreed with each other and gave the same rank, confirming $D = min(\, N_{L1} + \, N_{L2} - 1, N_{FOV}^2 \,)$. Also note that, unlike the $d_1 = d_2$ case for $K = 1$, different combinations of d_1, d_2 and d_3 values for $K = 2$ do not change the results or the upper bound of D, as also confirmed in Fig. 4.3.

3-Layer Case ($K = 3$):

For $K = 3$ case, we start with $\left(\boldsymbol{H}_{\boldsymbol{d_1}}^{T} \otimes \boldsymbol{H}_{\boldsymbol{d_4}}\right)\widehat{\boldsymbol{H}}_{\boldsymbol{d_{23}}}$. Then, we generate the matrix $\boldsymbol{H}'$ by discarding the columns of $\left(\boldsymbol{H}_{\boldsymbol{d_1}}^{T} \otimes \boldsymbol{H}_{\boldsymbol{d_4}}\right)\widehat{\boldsymbol{H}}_{\boldsymbol{d_{23}}}$ that correspond to the zero entries of $\boldsymbol{t_{123}}$. Here, the m^{th} column of $\boldsymbol{H}'$ is the vector that multiplies the coefficient $t_{1,i}t_{2,j}t_{3,k}$ for a certain (i, j, k) triplet. Hence, there is a one-to-one relationship between a given m and the pixel/neuron locations from the first, second and third diffractive layers, which are represented by $[q_{L1}, p_{L1}]$, $[q_{L2}, p_{L2}]$ and $[q_{L3}, p_{L3}]$, respectively. Similar to the $K = 1$ and $K = 2$ cases discussed in earlier sections, a given row, l, corresponds to a certain set of input (from N_i) and output (from N_o) pixels, $[q_i, p_i]$ and $[q_o, p_o]$, respectively. Accordingly, $\boldsymbol{H}'[l, m]$ can be written as:

$$\begin{aligned}\boldsymbol{H}'[l,m] = {} & h_{d1}(q_i - q_{L1}, p_i - p_{L1}) \cdot h_{d4}(q_o - q_{L3}, p_o - p_{L3}) \\ & \cdot \; h_{d2}(q_{L1} - q_{L2}, p_{L1} - p_{L2}) \cdot h_{d3}(q_{L2} - q_{L3}, p_{L2} - p_{L3})\end{aligned} \tag{4.12}$$

Then, we applied a coefficient and basis generation algorithm that is similar to Table 4.1, where we randomly select the diffractive layer and the neuron in each step of the algorithm to

obtain the resulting coefficients and the basis vectors. Then we converted the resulting vectors into their symbolic representations as discussed earlier for the $K = 1$ case and computed the rank of the resulting symbolic matrix. For $K = 3$ the selection order of the 1st, 2nd and 3rd diffractive layers in consecutive steps and the location/value of the chosen neuron at each step may affect the computed rank. Especially, when $N_{L1} + N_{L2} + N_{L3}$ is close to N_{FOV}^2, the probability of repeatedly achieving the upper-bound of the dimensionality of the solution space, i.e., $min(N_{L1} + N_{L2} + N_{L3} - 2, N_{FOV}^2)$, using random orders of selection decreases. In Fig. 4.4, we present the computed ranks for different K=3 diffractive network configurations; for each one of these configurations that we considered in our simulations, we obtained at least one random selection of the diffractive layers and neurons that attains full rank, numerically confirming $D = min(N_{L1} + N_{L2} + N_{L3} - 2, N_{FOV}^2)$.

The Upper Bound of the Dimensionality (D) of the Solution Space Reduces to $(N_{FOV}^2 + N_{FOV})/2$ when $d_1 = d_2$ for $K = 1$

For $K = 1$ and the special case of $d_1 = d_2 = d$, we can rewrite $\boldsymbol{H}'[l, m]$ given by Equation 4.8 as:

$$\boldsymbol{H}''[l, m] = h_d(q_i - q_{L1}, p_i - p_{L1})\, h_d(q_o - q_{L1}, p_o - p_{L1}). \tag{4.13}$$

To quantify the reduction in rank due to $d_1 = d_2$, among N_{FOV}^2 entries of $\boldsymbol{H}''[:, m]$, let us first consider the cases where $(q_i, p_i) \neq (q_o, p_o)$. For a given neuron or m, assuming that $(q_i, p_i) \neq (q_o, p_o)$, the number of different entries that can be produced by Equation 4.13

becomes $C\binom{N_{FOV}}{2}$, where $C(.)$ indicates the combination operation and $N_{FOV} = N_i = N_o$. Stated differently, since $h_{d1} = h_{d2}$ the order of the selections from (q_i, p_i) and (q_o, p_o) does not matter, making the selection defined by a combination operation, i.e., $C\binom{N_{FOV}}{2}$. In addition to these combinatorial entries, there are N_{FOV} additional entries that represent $(q_i, p_i) = (q_o, p_o)$. Therefore, the total number of unique entries in a column, $\boldsymbol{H}''[:, m]$, becomes:

$$C\binom{N_{FOV}}{2} + N_{FOV} = (N_{FOV}^2 + N_{FOV})/2. \tag{4.14}$$

This analysis proves that, for $K = 1$, the upper limit of the dimensionality (D) of the all-optical solution space for $d_1 = d_2$ reduces from N_{FOV}^2 to $(N_{FOV}^2 + N_{FOV})/2$ due to the fact that $h_{d1} = h_{d2} = h_d$ in Equation 4.13.

Note that, when $d_1 \neq d_2$, we have $h_{d1} \neq h_{d2}$, which directly implies that the combination operation in Equation S20 must be replaced with the permutation operation, $P(.)$, since the order of selections from (q_i, p_i) and (q_o, p_o) matters (see Equation S15). Therefore, when $d_1 \neq d_2$, Equation S20 is replaced with:

$$P\binom{N_{FOV}}{2} + N_{FOV} = N_{FOV}^2 \tag{4.15}$$

which confirms our analyses as well as the results reported in Fig. 4.2.

Numerical Analysis of Diffractive Networks

The previous section showed that the dimensionality of the all-optical solution space covered by K diffractive surfaces, forming an optical network positioned between an input and

output field-of-view, is determined by $\min\left(N_{FOV}^2, [\sum_{k=1}^{K} N_{Lk}] - (K-1)\right)$. However, this mathematical analysis does not shed light on the selection or optimization of the complex transmittance (or reflectance) values of each neuron of a diffractive network that is assigned for a given computational task. Here, we numerically investigate the function approximation power of multiple diffractive surfaces in the (N, K) space using image classification as a computational goal for the design of each diffractive network. Since N_{FOV} and N are large numbers in practice, an iterative optimization procedure based on error back-propagation and deep learning with a desired loss function was used to design diffractive networks and compare their performances as a function of (N, K).

For the first image classification task that was used as a test-bed, we formed nine different image data classes, where the input field-of-view (aperture) was randomly divided into nine different groups of pixels, each group defining one image class (Fig. 4.6a). Images of a given data class can have pixels only within the corresponding group, emitting light at arbitrary intensities towards the diffractive network. The computational task of each diffractive network is to blindly classify the input images from one of these nine different classes using *only nine large-area detectors* at the output field-of-view (Fig. 4.6b), where the classification decision is made based on the *maximum* of the optical signal collected by these nine detectors, each assigned to one particular image class. For deep-learning-based training of each diffractive network for this image classification task, we employed a cross-entropy loss function (see the Materials and Methods section).

Before we report the results of our analysis using a more standard image classification dataset such as CIFAR-10,[137] we initially selected this image classification problem defined in

Fig. 4.6 as it provides a well-defined linear transformation between the input and output fields-of-view. It also has various implications for designing new imaging systems with unique functionalities that cannot be covered by standard lens design principles.

Based on the diffractive network configuration and the image classification problem depicted in Fig. 4.6, we compared the training and blind testing accuracies provided by different diffractive networks composed of 1, 2 and 3 diffractive surfaces (each surface having N = 40K = 200×200 neurons) under different training and testing conditions (see Figs. 4.7-4.8). Our analysis also included the performance of a wider single-layer diffractive network with N = 122.5K > 3×40K neurons. For the training of these diffractive systems, we created two different training image sets (Tr_1 and Tr_2) to test the learning capabilities of different network architectures. In the first case, the training samples were selected such that approximately 1% of the point sources defining each image data class were simultaneously on and emitting light at various power levels.

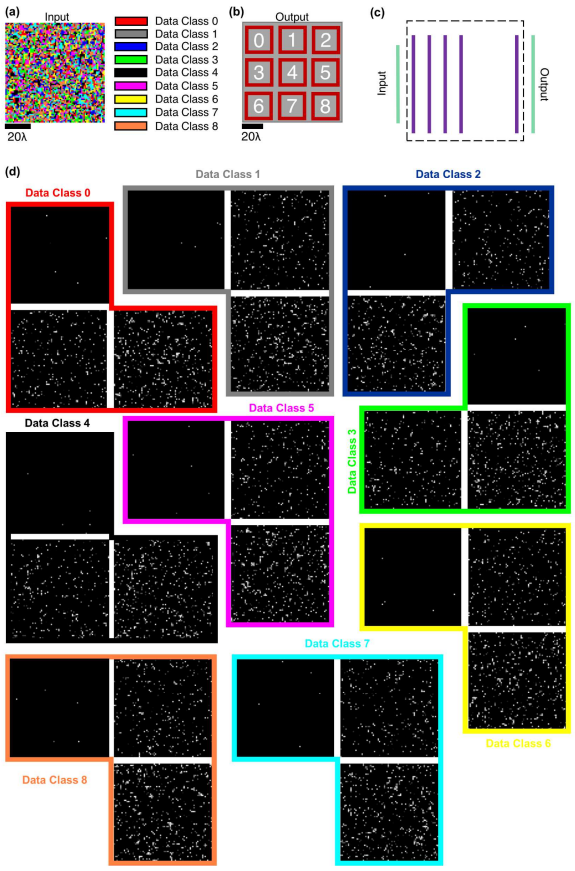

Fig. 4.6: Spatially-encoded image classification dataset. a A total of 9 image data classes are shown through color coding, defined inside the input field-of-view (80λ × 80λ). Each λ×λ area inside the field-of-view is randomly assigned to one image data class. An image belongs to a given data class, if and only if, all of its non-zero entries belong to the pixels that are assigned to that particular data class. b The layout of the 9 class detectors, positioned at the output plane. Each detector has an active area of 25λ × 25λ and for a given input image, the decision on class assignment is made based on the *maximum* optical signal among these 9 detectors. c Side view of the schematic of the diffractive network layers as well as the input and output fields-of-view. d Example images for 9 different data classes. Three samples for each image data class are illustrated here, randomly drawn from the 3 test datasets (Te_1, Te_{50}, and Te_{90}) that were used to quantify the blind inference accuracies of our diffractive network models (see Fig. 4.7).

For this training set, 200K images were created, forming Tr_1. In the second case, the training image dataset was constructed to include *only* a single point source (per image) located at different coordinates representing different data classes inside the input field-of-view, providing us with a total of 6.4K training images (which formed Tr_2). For the quantification of the blind testing accuracies of the trained diffractive models, three different test image datasets (never used during the training) were created, with each dataset containing 100K images. These three distinct test datasets (named Te_1, Te_{50} and Te_{90}) contain image samples that take contributions from 1% (Te_1), 50% (Te_{50}) and 90% (Te_{90}) of the points defining each image data class (see Fig. 4.6).

Figure 4.7 illustrates the blind classification accuracies achieved by the different diffractive network models that we trained. We see that as the number of diffractive surfaces in the network increases, the testing accuracies achieved by the final diffractive design improve significantly, meaning that the linear transformation space covered by the diffractive network expands with the addition of new trainable diffractive surfaces, in line with our former theoretical analysis. For instance, while a diffractive image classification network with a single phase-only (complex) modulation surface can achieve 24.48% (27.00%) for the test image set Te_1, the three-layer versions of the same architectures attain 85.2% (100.00%) blind testing accuracies, respectively (see Figs. 4.7a,b). Figure 4.8 shows the phase-only diffractive layers comprising the 1- and 3-layer diffractive optical networks that are compared in Fig. 4.7a; Fig. 4.8 also reports some exemplary test images selected from Te_1 and Te_{50}, along with the corresponding intensity distributions at the output planes of the diffractive networks. The comparison between two- and three-layer diffractive systems also indicates a similar conclusion

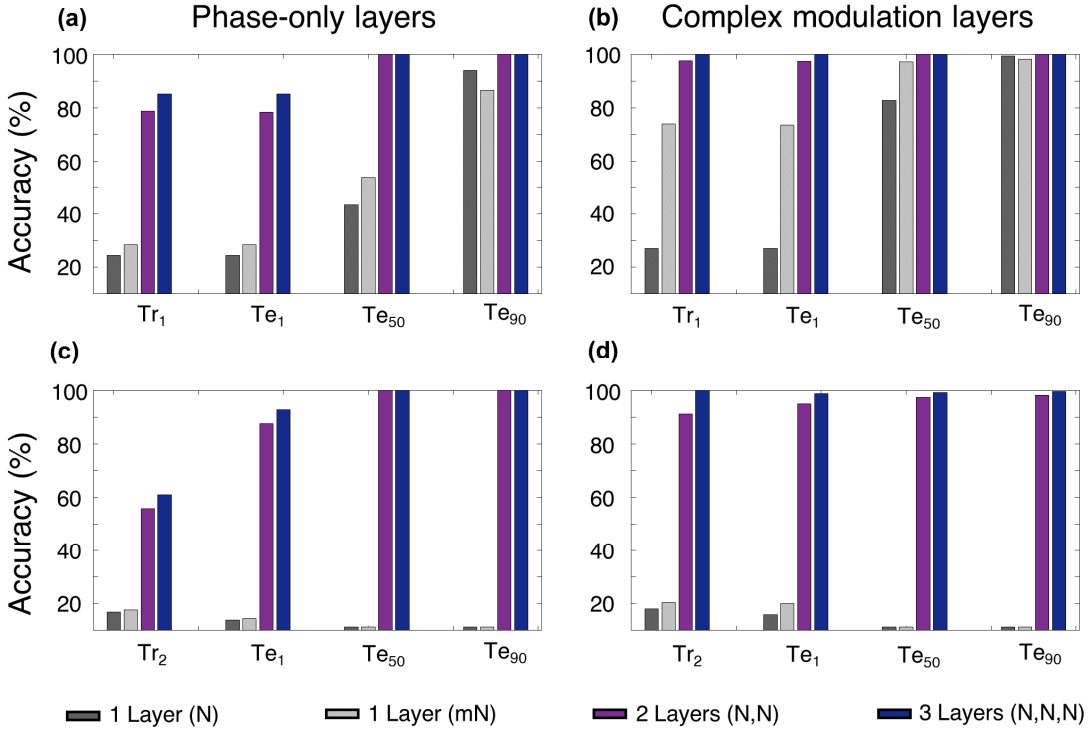

Fig. 4.7: Training and testing accuracy results for the diffractive surfaces that perform image classification (Figure 4.6). a The training and testing classification accuracies achieved by optical network designs comprised of diffractive surfaces that control only the phase of the incoming waves; the training image set is Tr_1 (200K images). b The training and testing classification accuracies achieved by optical network designs comprised of diffractive surfaces that can control both the phase and amplitude of the incoming waves; the training image set is Tr_1. c,d same as in a,b, respectively, except that the training image set is Tr_2 (6.4K images). N = 40K neurons, mN = 122.5K neurons, i.e., m>3.

for the test image set, Te_1. However, as we increase the number of point sources contributing to the test images, e.g., for the case of Te_{90}, the

blind testing classification accuracies of both the two- and three-layer networks saturate at nearly 100%, indicating that the solution space of the two-layer network already covers the optical transformation required to address this relatively easier image classification problem set by Te_{90}.

A direct comparison between the classification accuracies reported in Figs. 4.7a,c and Figs. 4.7b,d further reveals that the phase-only modulation constraint relatively limits the approximation power of the diffractive network since it places a restriction on the coefficients of the basis vectors, $\boldsymbol{h_{ij}}$. For example, when a two-layer, phase-only diffractive network is trained with Tr_1 and blindly tested with the images of Te_1, the training and testing accuracies are obtained as 78.72% and 78.44%, respectively. On the other hand, if the diffractive surfaces of the same network architectures have independent control of the transmission amplitude and phase value of each neuron of a given surface, the same training (Tr_1) and testing (Te_1) accuracy values increase to 97.68% and 97.39%, respectively.

As discussed in our earlier theoretical analysis, an alternative strategy to increase the all-optical processing capabilities of a diffractive network is to increase *N*, the number of neurons per diffractive surface. We also numerically investigated this scenario by training and testing another diffractive image classifier with a single surface that contains 122.5K neurons, i.e., it has more trainable neurons than the 3-layer diffractive designs reported in Fig. 4.7. As demonstrated in Fig. 4.7, although the performance of this larger/wider diffractive surface surpassed that of the previous, narrower/smaller 1-layer designs with 40K trainable neurons, its blind testing accuracy could not match the classification accuracies achieved by a 2-layer (2×40K neurons) network in both the phase-only and complex modulation cases. Despite using more trainable neurons than the 2-layer and 3-layer diffractive designs, the blind inference and generalization performance of this larger/wider diffractive surface is worse than that of the multi-surface diffractive designs. In fact, if we were to further increase the number of neurons in this single diffractive surface (further increasing the effective numerical aperture of the diffractive network), the inference performance gain due to these additional neurons that are farther away from the optical axis will

asymptotically go to zero since the corresponding k-vectors of these neurons carry a limited amount of optical power for the desired transformations targeted between the input and output fields-of-view.

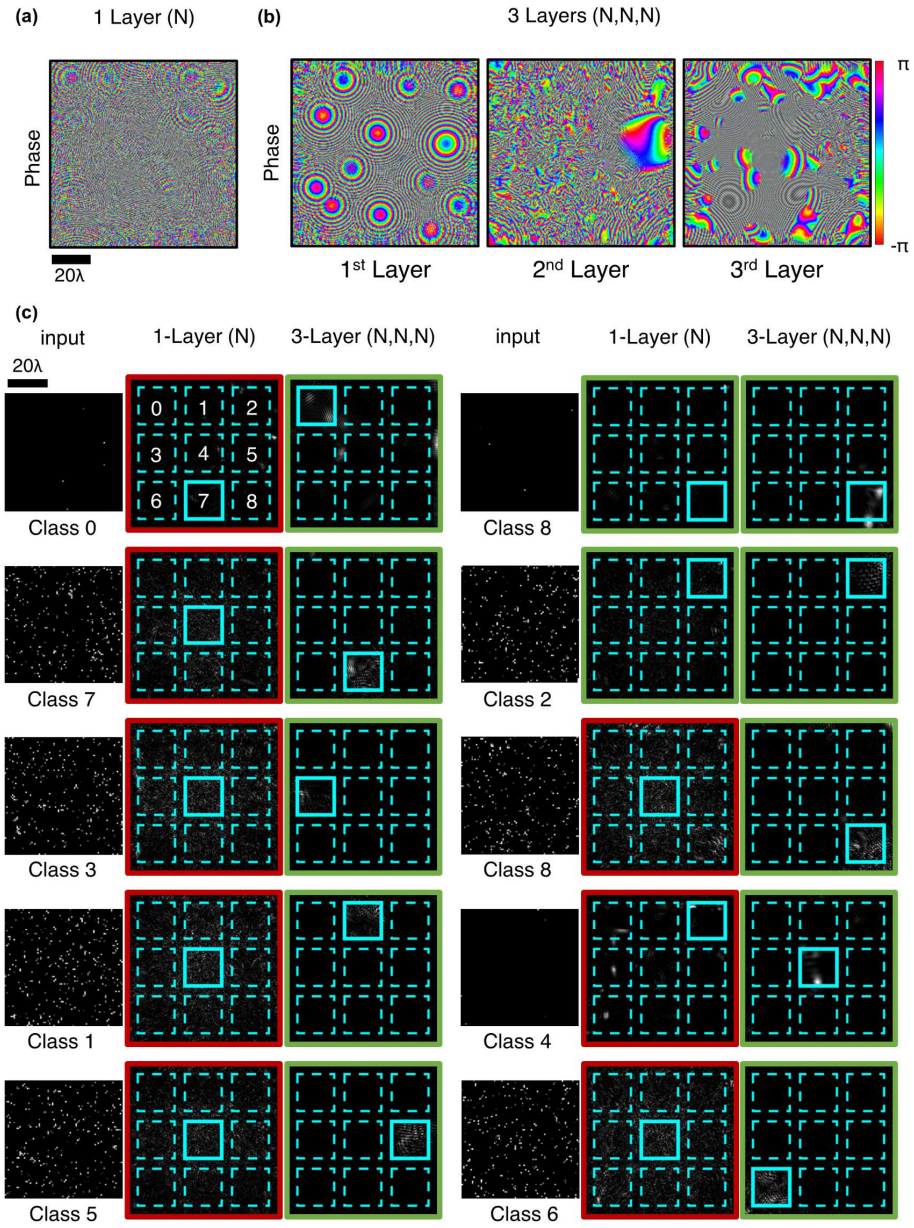

Fig. 4.8: 1- and 3-layer phase-only diffractive network designs and their input-output intensity profiles. a The phase profile of the single diffractive surface trained with Tr_1. b Same as in (a), except that there are 3 diffractive surfaces trained in the network design. c The output intensity distributions for the 1- and 3-layer diffractive networks shown in (a) and (b), respectively, for different input images, which were randomly selected from Te_1 and Te_{50}. A red (green) frame around the output intensity distribution indicates incorrect (correct) optical inference by the corresponding network. $N = 40K$.

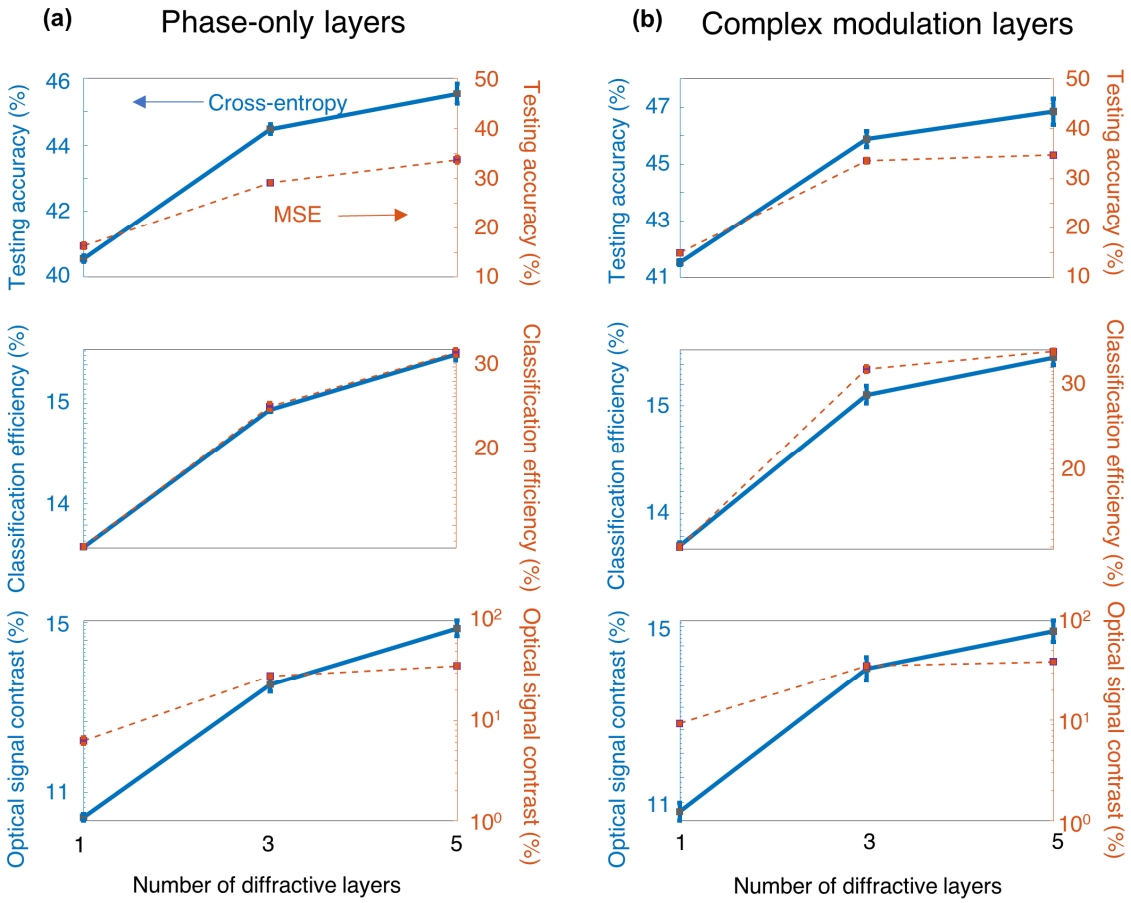

Fig. 4.9: The comparison of 1-, 3- and 5-layer diffractive networks trained for CIFAR-10 image classification, using MSE and cross-entropy loss functions. a Results for diffractive surfaces that modulate only the phase information of the incoming wave. b Results for diffractive surfaces that modulate both the phase and amplitude information of the incoming wave. The increase in the dimensionality of the all-optical solution space with additional diffractive surfaces of a network brings significant advantages in terms of generalization, blind testing accuracy, classification efficiency and optical signal contrast. The classification efficiency denotes the ratio of the optical power detected by the correct class detector with respect to the total detected optical power by all the class detectors at the output plane. Optical signal contrast refers to the normalized difference between the optical signals measured by the ground-truth (correct) detector and its strongest competitor detector at the output plane.

Another very important observation that one can make in Figs. 4.7c,d is that the performance improvements due to the increasing number of diffractive surfaces are much more pronounced for more challenging (i.e., limited) training image datasets, such as Tr_2. With a significantly smaller number of training images (6.4K images in Tr_2 as opposed to 200K images in Tr_1), multi-surface diffractive networks trained with Tr_2 successfully generalized to different test image datasets (Te_1, Te_{50} and Te_{90}) and efficiently learned the image classification problem at hand, whereas the single-surface diffractive networks (including the one with 122.5K trainable neurons per layer) almost entirely failed to generalize; see, e.g., Figs. 4.7c,d, the blind testing accuracy values for the diffractive models trained with Tr_2.

Next, we applied our analysis to a widely used, standard image classification dataset and investigated the performance of diffractive image classification networks comprised of one, three and five diffractive surfaces using the CIFAR-10 image dataset[137]. Unlike the previous image classification dataset (Fig. 4.6), the samples of CIFAR-10 contain images of physical objects, e.g., airplanes, birds, cats, dogs, etc., and CIFAR-10 has been widely used for quantifying the approximation power associated with various deep neural network architectures. Here, we assume that the CIFAR-10 images are encoded in the phase channel of the input field-of-view that is illuminated with a uniform plane wave. For deep-learning-based training of the diffractive classification networks, we adopted two different loss functions. The first loss function is based on the mean-squared-error (MSE), which essentially formulates the design of the all-optical object classification system as an image transformation/projection problem, and the second one is based on the cross-entropy loss, which is commonly used to solve the multi-class separation problems in the deep learning literature (refer to the Materials and Methods section for details).

The results of our analysis are summarized in Figs. 4.9a and 4.9b, which report the average blind inference accuracies along with the corresponding standard deviations observed over the testing of three different diffractive network models trained independently to classify the CIFAR-10 test images using phase-only and complex-valued diffractive surfaces, respectively. The 1-, 3-, and 5-layer phase-only (complex-valued) diffractive network architectures can attain blind classification accuracies of $40.55 \mp 0.10\%$ ($41.52 \mp 0.09\%$), $44.47 \mp 0.14\%$ ($45.88 \mp 0.28\%$) and $45.53 \mp 0.30\%$ ($46.84 \mp 0.46\%$), respectively, when they are trained based on the cross-entropy loss detailed in the Materials and Methods section. On the other hand, with the use of the MSE loss, these classification accuracies are reduced to $16.25 \mp 0.48\%$ ($14.92 \mp 0.26\%$), $29.08 \mp 0.14\%$ ($33.52 \mp 0.40\%$) and $33.67 \mp 0.57\%$ ($34.69 \mp 0.11\%$), respectively. In agreement with the conclusions of our previous results and the presented theoretical analysis, the blind testing accuracies achieved by the all-optical diffractive classifiers improve with increasing number of diffractive layers, K, independent of the loss function used and the modulation constraints imposed on the trained surfaces (see Fig. 4.9).

Different from electronic neural networks, however, diffractive networks are physical machine learning platforms with their own optical hardware; hence, practical design merits such as the signal-to-noise ratio (SNR) and the contrast-to-noise ratio (CNR) should also be considered, as these features can be critical for the success of these networks in various applications. Therefore, in addition to the blind testing accuracies, the performance evaluation and comparison of these all-optical diffractive classification systems involve two additional metrics that are analogous to the SNR and CNR. The first is the *classification efficiency*, which we define as the ratio of the optical signal collected by the target, ground-truth class detector, I_{gt}, with respect to the total power collected by all class detectors located at the output plane. The

second performance metric refers to the normalized difference between the optical signals measured by the ground-truth/correct detector, I_{gt}, and its strongest competitor, I_{sc}, i.e., $(I_{gt} - I_{sc}) / I_{gt}$; this optical signal contrast metric is, in general, important since the relative level of detection noise with respect to this difference is critical for translating the accuracies achieved by the numerical forward models to the performance of the physically fabricated diffractive networks. Figure 4.9 reveals that the improvements observed in the blind testing accuracies as a function of the number of diffractive surfaces also apply to these two important diffractive network performance metrics, resulting from the increased dimensionality of the all-optical solution space of the diffractive network with increasing *K*. For instance, the diffractive network models presented in Fig. 4.9b, trained with the cross-entropy (or MSE) loss function, provide *classification efficiencies* of 13.72∓0.03% (13.98∓0.12%), 15.10∓0.08% (31.74∓0.41%) and 15.46 ∓ 0.08% (34.43 ∓ 0.28%) using complex-valued 1-, 3- and 5-layers, respectively. Furthermore, the *optical signal contrast* attained by the same diffractive network designs can be calculated as 10.83∓0.17% (9.25∓0.13%), 13.92∓0.28% (35.23∓1.02%) and 14.88∓0.28% (38.67 ∓ 0.13%), respectively. Similar improvements are also observed for the phase-only diffractive optical network models that are reported in Fig. 4.9a. These results indicate that the increased dimensionality of the solution space with increasing *K* improves the inference capacity as well as the robustness of the diffractive network models by enhancing their optical efficiency and signal contrast.

Apart from the results and analyses reported in this section, the depth advantage of diffractive networks has been empirically shown in the literature for some other applications and datasets, such as, e.g., image classification[77,78] and optical spectral filter design[80].

4.3 Discussion

In a diffractive optical design problem, it is not guaranteed that the diffractive surface profiles will converge to the optimum solution for a given (N, K). Furthermore, for most applications of interest such as image classification, the optimum transformation matrix that the diffractive surfaces need to approximate is unknown; for example, what defines all the images of cats vs. dogs (such as in CIFAR-10 image dataset) is not known analytically to create a target transformation. Nonetheless, it can be argued that as the dimensionality of the all-optical solution space, and thus the approximation power of the diffractive surfaces increases, the probability of converging to a solution satisfying the desired design criteria also increases. In other words, even if the optimization of the diffractive surfaces gets stuck in a local minimum, which is practically always the case, there is a greater chance that this state will be closer to the globally optimal solution(s) for deeper diffractive networks with multiple trainable surfaces.

Although not considered in our analysis so far, an interesting future direction to investigate is the case when the axial distance between two successive diffractive surfaces is made much smaller than the wavelength of light, i.e., $d << \lambda$. In this case, all the evanescent waves and the surface modes of each diffractive layer would need to be carefully taken into account to analyze the all-optical processing capabilities of the resulting diffractive network. This would significantly increase the space-bandwidth product of the optical processor as the effective neuron size per diffractive surface/layer can be deeply subwavelength if the near-field is taken into account. Furthermore, due to the presence of near-field coupling between diffractive surfaces/layers, the effective transmission or reflection coefficient of each neuron of a surface will no longer be an independent parameter as it will depend on the configuration/design of the other surfaces. If all of these near-field related coupling effects are carefully taken into account

during the design of a diffractive optical network with $d << \lambda$, it can significantly enrich the solution space of multi-layer coherent optical processors, assuming that surface fabrication resolution and the signal-to-noise ratio as well as the dynamic range at the detector plane are all sufficient. Despite the theoretical richness of near-field-based diffractive optical networks, the design and implementation of such systems bring substantial challenges in terms of their 3D fabrication and alignment as well as the accuracy of the computational modelling of the associated physics within the diffractive network, including multiple reflections and boundary conditions. While various electromagnetic wave solvers can handle the numerical analysis of near-field diffractive systems, practical aspects of a fabricated near-field diffractive neural network will present various sources of imperfections and errors that might force the physical forward model to significantly deviate from numerical simulations.

In summary, we presented a theoretical analysis on the information processing capacity and function approximation power of diffractive surfaces that can compute a given task using temporally and spatially coherent light. In our analysis, we assumed that the polarization state of the propagating light is preserved by the optical modulation on the diffractive surfaces and the axial distance between successive layers is kept large enough to ensure that the near-field coupling and related effects can be ignored in the optical forward model. Based on these assumptions, our analysis shows that the dimensionality of the all-optical solution space provided by multi-layer diffractive networks expands linearly as a function of the number of trainable surfaces, K, until it reaches the limit defined by the target input and output fields-of-view, i.e., $\min\left(N_i N_o, [\sum_{k=1}^{K} N_{Lk}] - (K-1)\right)$ as depicted in Equation 4.7. To numerically validate these conclusions, we adopted a deep learning-based training strategy to design diffractive image classification systems for two distinct datasets (Figs. 4.6-4.9) and investigated

their performance in terms of blind inference accuracy, learning and generalization performance, classification efficiency and optical signal contrast, confirming the depth advantages provided by multiple diffractive surfaces compared to a single diffractive layer.

These results and conclusions, along with the underlying analyses, broadly cover various types of diffractive surfaces including e.g., metamaterials/metasurfaces, nanoantenna arrays, plasmonics and flat optics based designer surfaces. We believe that the deeply subwavelength design features of e.g., diffractive metasurfaces can open up new avenues in the design of coherent optical processers by enabling independent control over the amplitude and phase modulation of neurons of a diffractive layer, also providing unique opportunities to engineer the material dispersion properties as needed for a given computational task.

4.4 Materials and Methods

Coefficient and basis vector generation for an optical network formed by two diffractive surfaces

Here we present the details of the coefficient and basis vector generation algorithm for a network having two diffractive surfaces with the neurons (N_{L1}, N_{L2}) to show that it is capable of forming a vectorized transformation matrix in $N_{L1}+N_{L2}-1$ dimensional subspace of an $N_i N_o$-dimensional complex-valued vector space. The algorithm depends on consuming the transmittance values from the first or the second diffractive layer, i.e., $\boldsymbol{T_1}$ or $\boldsymbol{T_2}$, at each step after its initialization. Choosing a random neuron from $\boldsymbol{T_1}$ or $\boldsymbol{T_2}$ is followed by forming a new basis vector. The chosen neuron becomes the coefficient of this new basis vector which is generated by using the previously chosen transmittance values and appropriate vectors from $\boldsymbol{h_{ij}}$ (Equation 4.5). The algorithm continues until all the transmittance values are assigned to an arbitrary complex-valued coefficient and using all the vectors of $\boldsymbol{h_{ij}}$ in forming the basis vectors.

In Table 4.1, a pseudo-code of the algorithm is also presented. In this table, $C_{1,k}$ and $C_{2,k}$ represent the sets of the transmittance values that include $t_{1,i}$ and $t_{2,j}$ which were not chosen before (at the time step k), from the first and second diffractive surfaces, respectively. Also, $c_k = t_{1,i}$ in Step 7 and $c_k = t_{2,j}$ in Step 10 are the complex-valued coefficients that can be independently determined. Similarly $\boldsymbol{b_k} = \sum_{t_{2,j} \notin C_{2,k}} t_{2,j} \boldsymbol{h_{ij}}$ and $\boldsymbol{b_k} = \sum_{t_{1,i} \notin C_{1,k}} t_{1,i} \boldsymbol{h_{ij}}$ are the generated basis vectors at each step, where $t_{1,i} \notin C_{1,k}$ and $t_{2,j} \notin C_{2,k}$ represent the sets of coefficients which are chosen before. The basis vectors in Step 7 and Step 10 are formed through the linear combinations of some $\boldsymbol{h_{ij}}$ vectors. Since the total number of vectors generated by this method is $N_{L1}+N_{L2}-1 < N_i N_o$ (discussed in the following paragraph), it is guaranteed that the generated $\boldsymbol{b_k}$ at each step k is independent from the previously generated basis vectors.

By examining the algorithm in Table 4.1, it is straightforward to show that the total number of generated basis vectors is $N_{L1}+N_{L2}-1$. That is, at each time step k, only one coefficient is chosen and only one basis vector is created. Since there are $N_{L1}+N_{L2}$ many transmittance values where two of them are chosen together in Step 1, the total number of time steps (coefficient and basis vectors) become $N_{L1}+N_{L2}-1$. On the other hand, showing that all the $N_{L1}N_{L2}$-many $\boldsymbol{h_{ij}}$ vectors are used in the algorithm requires further analysis. Without loss of generality, let $\boldsymbol{T_1}$ be chosen n_1 times starting from the time step $k = 2$ and then $\boldsymbol{T_2}$ is chosen n_2 times. Similarly, $\boldsymbol{T_1}$ and $\boldsymbol{T_2}$ are chosen n_3 and n_4 times in the following cycles, respectively. Then, this pattern follows until all the $N_{L1}+N_{L2}$ many transmittance values are consumed. Here we show the partition of the selection of the transmittance values from $\boldsymbol{T_1}$ and $\boldsymbol{T_2}$ for each time step k into s many chunks, i.e.:

$$k = \left\{ \underbrace{2,3,\ldots}_{n_1}, \underbrace{\ldots}_{n_2}, \underbrace{\ldots}_{n_3}, \underbrace{\ldots}_{n_4}, \ldots, \underbrace{\ldots N_{L1}+N_{L2}-2, N_{L1}+N_{L2}-1}_{n_s} \right\} \tag{4.16}$$

1	*Randomly choose $t_{1,i}$ from the set $C_{1,1}$ and $t_{2,j}$ from the set $C_{2,1}$, and assign desired values to the chosen $t_{1,i}$ and $t_{2,j}$*

In order show that, $N_{L1}N_{L2}$-many $\boldsymbol{h_{ij}}$ vectors are used in the algorithm regardless of the values of *s* and n_i, we first define

$$p_i = n_i + p_{i-2} \text{ for even values of i} \geq 2$$

$$q_i = n_i + q_{i-2} \text{ for odd values of i} \geq 1$$

where $p_0 = 0$ and $q_{-1} = 1$. Based on this, the total number of consumed basis vectors

$$\begin{aligned} n_h = 1 &+ \sum_{k=2}^{q_1} 1 + \sum_{k=q_1+1}^{p_2+q_1} q_1 + \sum_{k=p_2+q_1+1}^{q_3+p_2} (p_2+1) + \sum_{k=q_3+p_2+1}^{p_4+q_3} q_3 \\ &+ \sum_{k=p_4+q_3+1}^{q_5+p_4} (p_4+1) + \sum_{k=q_5+p_4+1}^{p_6+q_5} q_5 + \sum_{k=p_6+q_5+1}^{q_7+p_6} (p_6+1) \\ &+ \quad \cdots \\ &+ \sum_{k=p_{s-2}+q_{s-3}+1}^{N_{L1}+p_{s-2}} (p_{s-2}+1) + \sum_{k=N_{L1}+p_{s-2}+1}^{N_{L1}+N_{L2}-1} N_{L1} \end{aligned} \tag{4.17}$$

inside each summation in Table 4.1 (Steps 7 and 10) can be written as:

where each summation gives the number of the consumed $\boldsymbol{h_{ij}}$ vectors in the corresponding chunk. Please note that, based on the partition given by Equation 4.17, q_{s-1} and p_s become equal to N_{L1} and $N_{L2}-1$, respectively. One can show, by carrying out this summation, that all the terms except $N_{L1}N_{L2}$ cancel each other, and therefore $n_h = N_{L1}N_{L2}$

2	$c_1\boldsymbol{b_1} = t_{1,i}t_{2,j}\boldsymbol{h_{ij}}$
3	*k=2*
4	*Randomly choose* $\boldsymbol{T_1}$ *or* $\boldsymbol{T_2}$ *if* $C_{1,k} \neq \emptyset$ *and* $C_{2,k} \neq \emptyset$ *Choose* $\boldsymbol{T_1}$ *if* $C_{1,k} \neq \emptyset$ *and* $C_{2,k} = \emptyset$ *Choose* $\boldsymbol{T_2}$ *if* $C_{1,k} = \emptyset$ *and* $C_{2,k} \neq \emptyset$
5	*If* $\boldsymbol{T_1}$ *is chosen in Step 4:*
6	*Randomly choose* $t_{1,i}$ *from the set* $C_{1,k}$*, and assign a desired value to the chosen* $t_{1,i}$
7	$c_k\boldsymbol{b_k} = t_{1,i}\left(\sum_{t_{2,j}\notin C_{2,k}} t_{2,j}\boldsymbol{h_{ij}}\right)$
8	*else:*
9	*Randomly choose* $t_{2,j}$ *from the set* $C_{2,k}$*, and assign a desired value to the chosen* $t_{2,j}$
10	$c_k\boldsymbol{b_k} = t_{2,j}\left(\sum_{t_{1,i}\notin C_{1,k}} t_{1,i}\boldsymbol{h_{ij}}\right)$
11	*k = k+1*
12	*If* $C_{1,k} \neq \emptyset$ *or* $C_{2,k} \neq \emptyset$*:*
13	*Return to Step 4*
14	*else:*
15	*Exit*

Table 4.1 Coefficient (c_k) and basis vector (b_k) generation algorithm pseudo-code for an optical network that has two diffractive surfaces.

Step	Choice from $\boldsymbol{T_1}$	Choice from $\boldsymbol{T_2}$	Choice from $\boldsymbol{T_3}$	Resulting Coefficient and Basis Vector
1	$t_{1,1}$	$t_{2,1}$ (fixed)	$t_{3,1}$ (fixed)	$c_1\boldsymbol{b_1} = t_{1,1}(t_{2,1}t_{3,1}\boldsymbol{h_{111}})$
2	-	-	$t_{3,2}$	$c_2\boldsymbol{b_2} = t_{3,2}(t_{1,1}t_{2,1}\boldsymbol{h_{112}})$
3	-	$t_{2,2}$	-	$c_3\boldsymbol{b_3} = t_{2,2}(t_{1,1}t_{3,1}\boldsymbol{h_{121}} + t_{1,1}t_{3,2}\boldsymbol{h_{122}})$
4	$t_{1,2}$	-	-	$c_4\boldsymbol{b_4} = t_{1,2}(t_{2,1}t_{3,1}\boldsymbol{h_{211}} + t_{2,1}t_{3,2}\boldsymbol{h_{212}} + t_{2,2}t_{3,1}\boldsymbol{h_{221}} + t_{2,2}t_{3,2}\boldsymbol{h_{222}})$
5	-	-	$t_{3,3}$	$c_5\boldsymbol{b_5} = t_{3,3}(t_{1,1}t_{2,1}\boldsymbol{h_{113}} + t_{1,1}t_{2,2}\boldsymbol{h_{123}} + t_{1,2}t_{2,1}\boldsymbol{h_{213}} + t_{1,2}t_{2,2}\boldsymbol{h_{223}})$
6	-	$t_{2,3}$	-	$c_6\boldsymbol{b_6} = t_{2,3}(t_{1,1}t_{3,1}\boldsymbol{h_{131}} + t_{1,1}t_{3,2}\boldsymbol{h_{132}} + t_{1,1}t_{3,3}\boldsymbol{h_{133}} + t_{1,2}t_{3,1}\boldsymbol{h_{231}} + +t_{1,2}t_{3,2}\boldsymbol{h_{232}} + t_{1,2}t_{3,3}\boldsymbol{h_{233}})$
7	$t_{1,3}$	-	-	$c_7\boldsymbol{b_7} = t_{1,3}(t_{2,1}t_{3,1}\boldsymbol{h_{311}} + t_{2,1}t_{3,2}\boldsymbol{h_{312}} + t_{2,1}t_{3,3}\boldsymbol{h_{313}} + t_{2,2}t_{3,1}\boldsymbol{h_{321}} + t_{2,2}t_{3,2}\boldsymbol{h_{322}} + t_{2,2}t_{3,3}\boldsymbol{h_{323}} + t_{2,3}t_{3,1}\boldsymbol{h_{331}} + t_{2,3}t_{3,2}\boldsymbol{h_{332}} + t_{2,3}t_{3,3}\boldsymbol{h_{333}})$
⋮	⋮	⋮	⋮	⋮
3n-4	-	-	$t_{3,n}$	$c_{3n-4}\boldsymbol{b_{3n-4}} = t_{3,n}\left(\sum_{i=1}^{n-1}\sum_{j=1}^{n-1} t_{1,i}t_{2,j}\boldsymbol{h_{ijn}}\right)$
3n-3	-	$t_{2,n}$	-	$c_{3n-3}\boldsymbol{b_{3n-3}} = t_{2,n}\left(\sum_{i=1}^{n-1}\sum_{k=1}^{n} t_{1,i}t_{3,k}\boldsymbol{h_{ink}}\right)$
3n-2	$t_{1,n}$	-	-	$c_{3n-2}\boldsymbol{b_{3n-2}} = t_{1,n}\left(\sum_{j=1}^{n}\sum_{k=1}^{n} t_{2,j}t_{3,k}\boldsymbol{h_{njk}}\right)$
⋮	⋮	⋮	⋮	⋮
3N-4	-	-	$t_{3,N}$	$c_{3N-4}\boldsymbol{b_{3N-4}} = t_{3,N}\left(\sum_{i=1}^{N-1}\sum_{j=1}^{N-1} t_{1,i}t_{2,j}\boldsymbol{h_{ijN}}\right)$
3N-3	-	$t_{2,N}$	-	$c_{3N-3}\boldsymbol{b_{3N-3}} = t_{2,N}\left(\sum_{i=1}^{N-1}\sum_{k=1}^{N} t_{1,i}t_{3,k}\boldsymbol{h_{iNk}}\right)$
3N-2	$t_{1,N}$	-	-	$c_{3N-2}\boldsymbol{b_{3N-2}} = t_{1,N}\left(\sum_{j=1}^{N}\sum_{k=1}^{N} t_{2,j}t_{3,k}\boldsymbol{h_{Njk}}\right)$

Table 4.2 Coefficient and basis vector generation algorithm for a 3-layered diffractive network (K = 3) when $\mathrm{N_{L1}} = \mathrm{N_{L2}} = \mathrm{N_{L3}} = \mathrm{N_i} = \mathrm{N_o} = \mathrm{N}$.

Step	Choice from $\boldsymbol{T_1}$	Choice from $\boldsymbol{T_2}$	Resulting Coefficient and Basis Vector
1	$t_{1,1}$	$t_{2,1}$ (fixed)	$c_1\boldsymbol{b_1} = t_{1,1}(t_{2,1}\boldsymbol{h_{11}})$
2	-	$t_{2,2}$	$c_2\boldsymbol{b_2} = t_{2,2}(t_{1,1}\boldsymbol{h_{12}})$
3	$t_{1,2}$	-	$c_{i+1}\boldsymbol{b_{i+1}} = t_{1,i}(t_{2,1}\boldsymbol{h_{i1}} + t_{2,2}\boldsymbol{h_{i2}})$, for $i \in \{2,3,\dots,K\}$
4	$t_{1,3}$	-	
⋮	⋮	⋮	
K+1	$t_{1,K}$	-	
K+2	-	$t_{2,3}$	$c_{K+2}\boldsymbol{b}_{K+2} = t_{2,3}\left(\sum_{i=2}^{K} t_{1,i}\boldsymbol{h_{i1}}\right)$
K+3	$t_{1,K+1}$	-	$c_{i+2}\boldsymbol{b_{i+2}} = t_{1,i}(t_{2,1}\boldsymbol{h_{i1}} + t_{2,2}\boldsymbol{h_{i2}} + t_{2,3}\boldsymbol{h_{i3}})$, for $i \in \{K+1, K+2, \dots, 2K-1\}$
K+4	$t_{1,K+2}$	-	
⋮	⋮	⋮	
2K+1	$t_{1,2K-1}$	-	
⋮	⋮	⋮	⋮
qK+2	-	$t_{2,q+2}$	$c_{qK+2}\boldsymbol{b}_{qK+2} = t_{2,q+2}\left(\sum_{i=1}^{qK-(q-1)} t_{1,i}\boldsymbol{h}_{i(q+2)}\right)$
qK+3	$t_{1,q(K-1)+2}$	-	$c_{i+1}\boldsymbol{b_{i+1}} = t_{1,i-q}\left(\sum_{j=1}^{q+2} t_{2,j}\boldsymbol{h}_{(i-q)j}\right)$, for $i \in \{qK+2, qK+3, \dots, qK+K\}$
qK+4	$t_{1,q(K-1)+3}$	-	
⋮	⋮	⋮	
(q+1)K+1	$t_{1,(q+1)K-q}$	-	
⋮	⋮	⋮	⋮
(N-2)K+2	-	$t_{2,N}$	$c_{(N-2)K+2}\boldsymbol{b}_{(N-2)K+2} = t_{2,N}\left(\sum_{i=1}^{(N-2)K-(N-1)} t_{1,i}\boldsymbol{h_{iN}}\right)$
(N-2)K+3	$t_{1,(N-2)(K-1)+2}$	-	$c_{i+N-1}\boldsymbol{b_{i+N-1}} = t_{1,i}\left(\sum_{j=1}^{N} t_{2,j}\boldsymbol{h_{ij}}\right)$, for $i \in \{(N-2)(K-1)+2, (N-2)(K-1)+3, \dots, (N-1)K-(N-2)\}$
(N-2)K+4	$t_{1,(N-2)(K-1)+3}$	-	
⋮	⋮	⋮	
KN-(K-1)	$t_{1,(N-1)K-(N-2)}$	-	

Table 4.3 Coefficient and basis vector generation algorithm for a K-layered diffractive network when $N_{L1} = N_{L2} = N_{L3} = N_i = N_o = N$.

showing that all the $N_{L1}N_{L2}$-many $\boldsymbol{h_{ij}}$ vectors are used in the algorithm. Here we assumed that the transmittance values from the first diffractive layer are consumed first. However, even if it were assumed that the transmittance values from the second diffractive layer is consumed first, the result would not change.

Optical Forward Model

In a coherent optical processor composed of diffractive surfaces, the optical transformation between a given pair of input/output fields-of-view is established through the modulation of light by a series of diffractive surfaces which we modeled as two-dimensional, thin, multiplicative elements. According to our formulation, the complex-valued transmittance of a diffractive surface, *k*, is defined as;

$$t(x, y, z_k) = a(x, y) \exp\big(j2\pi\phi(x, y)\big) \tag{4.18}$$

where $a(x, y)$ and $\phi(x, y)$ denote the trainable amplitude and the phase modulation functions of diffractive layer *k*. The values of $a(x, y)$, in general, lie in the interval (0, 1), i.e., there is no optical gain over these surfaces, and the dynamic range of the phase modulation is between (0, 2π). In the case of phase-only modulation restriction, however, $a(x, y)$ is kept as 1 (non-trainable) for all the neurons. The parameter z_k defines the axial location of the diffractive layer *k* between the input field-of-view at $z = 0$ and the output plane. Based on these assumptions, the Rayleigh-Sommerfeld formulation expresses the light diffraction by modelling each diffractive unit on layer *k* at $\big(x_q, y_q, z_k\big)$ as the source of a secondary wave:

$$w_q^k(x, y, z) = \frac{z - z_k}{r^2}\left(\frac{1}{2\pi r} + \frac{1}{j\lambda}\right)\exp\left(\frac{j2\pi r}{\lambda}\right) \tag{4.19}$$

where $r = \sqrt{(x - x_q)^2 + (y - x_q)^2 + (z - z_k)^2}$. Combining Equations 4.18 and 4.19, we can write the light field exiting the q^{th} diffractive unit of layer $k+1$ as:

$$u_q^{k+1}(x, y, z) = t(x_q, y_q, z_{k+1}) w_q^{k+1}(x, y, z) \sum_{p \in S_k} u_p^k(x_q, y_q, z_{k+1}) \tag{4.20}$$

where S_k denotes the set of diffractive units of layer k. From Equation 4.20, the complex wave field at the output plane can be written as:

$$u^{K+1}(x, y, z) = \sum_{q \in S_K} \left[t(x_q, y_q, z_K) w_q^K(x, y, z) \sum_{p \in S_{K-1}} u_p^{K-1}(x_q, y_q, z_K) \right] \tag{4.21}$$

where the optical field immediately after the object is assumed to be $u^0(x, y, z)$. In Equation 4.21, S_K and S_{K-1} denote the set of features at the K^{th} and $(K\text{-}1)^{\text{th}}$ diffractive layers, respectively.

Image classification datasets and diffractive network parameters

There are a total of nine image classes in the dataset defined in Fig. 4.6, corresponding to nine different sets of coordinates inside the input field-of-view, which covers a region of 80λ × 80λ. Each point source lies inside a region of λ × λ, resulting in 6.4K coordinates, divided into nine image classes. Nine classification detectors were placed at the output plane, each representing a data class, as depicted in Fig. 4.6b. The sensitive area of each detector was set to 25λ × 25λ. In this design, the classification decision was made based on the *maximum* of the optical signal collected by these nine detectors. According to our system architecture, the image in the field-of-view and the class detectors at the output plane were connected through diffractive surfaces of size 100λ × 100λ, and for the multi-layer ($K > 1$) configurations, the axial distance, d,

between two successive diffractive surfaces was taken as 40λ. With a neuron size of λ/2, we obtained N = 40K (200×200), N_i = 25.6K (160×160) and N_o = 22.5K (9×50×50).

For the classification of the CIFAR-10 image dataset, the size of the diffractive surfaces was taken to be approximately 106.6λ × 106.6λ, and the edge length of the input field-of-view containing the input image was set to be ~53.3λ in both lateral directions. Unlike the amplitude encoded images of the previous dataset (Fig. 4.6), the information of the CIFAR-10 images was encoded in the phase channel of the input field, i.e., a given input image was assumed to define a phase-only object with the grey levels corresponding to the delays experienced by the incident wavefront within the range [0, λ). To form the phase-only object inputs based on the CIFAR-10 dataset, we converted the RGB samples to greyscale by computing their YCrCb representations. Then, unsigned 8-bit integer values in the Y channel were converted into float32 values and normalized to the range [0, 1]. These normalized greyscale images were then mapped to phase values between [0, 2π). The original CIFAR-10 dataset[137] has 50K training and 10K test images. In the diffractive optical network designs presented here, we used all 50K and 10K images during the training and testing stages, respectively. Therefore, the blind classification accuracy, efficiency and optical signal contrast values depicted in Fig. 4.9 were computed over the entire 10K test set.

The responsivity of the 10 class detectors placed at the output plane (each representing one CIFAR-10 data class, e.g., automobile, ship, truck, etc.) was assumed to be identical and uniform over an area of 6.4λ × 6.4λ. The axial distance between two successive diffractive

surfaces in the design was assumed to be 40λ. Similarly, the input and output fields-of-view were placed 40λ away from the first and last diffractive layers, respectively.

Loss functions and training details

For a given dataset with C classes, one way of designing an all-optical diffractive classification network is to place C class detectors at the output plane, establishing a one-to-one correspondence between data classes and the opto-electronic detectors. Accordingly, the training of these systems aims to find/optimize the diffractive surfaces that can route most of the input photons, thus the optical signal power, to the corresponding detector representing the data class of a given input object.

The first loss function that we used for the training of diffractive optical networks is the cross-entropy loss, which is frequently used in machine learning for multi-class image classification. This loss function acts on the optical intensities collected by the class detectors at the output plane and is defined as:

$$\mathcal{L} = -\sum_{c \,\epsilon\, C} g_c \log(\sigma_c) \tag{4.22}$$

where g_c and σ_c denote the entry in the one-hot label vector and the class score of class c, respectively. The class score σ_c, on the other hand, is defined as a function of the normalized optical signals, $\boldsymbol{I'}$;

$$\sigma_c = \frac{\exp(I'_c)}{\sum_{c \,\epsilon\, C} \exp(I'_c)} \tag{4.23}$$

Equation 4.23 is the well-known softmax function. The normalized optical signals $\boldsymbol{I}'$ are defined as $\frac{\boldsymbol{I}}{\max\{\boldsymbol{I}\}} \times T$, where $\boldsymbol{I}$ is the vector of the detected optical signals for each class detector and T is a constant parameter that induces a virtual contrast, helping to increase the efficacy of training.

Alternatively, the all-optical classification design achieved using a diffractive network can be cast as a coherent image projection problem by defining a ground-truth spatial intensity profile at the output plane for each data class and an associated loss function that acts over the synthesized optical signals at the output plane. Accordingly, the mean-squared-error (MSE) loss function used in Fig. 4.9 computes the difference between a ground-truth intensity profile, $I_g^c(x,y)$, devised for class c and the intensity of the complex wave field at the output plane, i.e., $|u^{K+1}(x,y)|^2$. We defined $I_g^c(x,y)$ as:

$$I_g^c(x,y) = \begin{cases} 1 & if\ x \in D_x^c\ and\ y \in D_y^c \\ 0 & otherwise \end{cases} \tag{4.24}$$

where D_x^c and D_y^c represent the sensitive/active area of the class detector corresponding to class c. The related MSE loss function, $\mathcal{L}_{mse}$, can then be defined as:

$$\mathcal{L}_{mse} = \int\int \left||u^{K+1}(x,y)|^2 - I_g^c(x,y)\right|^2 dxdy \tag{4.25}$$

All network models used in this work were trained using Python (v3.6.5) and TensorFlow (v1.15.0, Google Inc.). We selected the Adam[106] optimizer during the training of all the models, and its parameters were taken as the default values used in TensorFlow and kept identical in each model. The learning rate of the diffractive optical networks was set to 0.001.

Chapter 5 All-optical Synthesis of An Arbitrary Linear Transformation Using Diffractive Surfaces

Parts of this chapter have previously been published in O. Kulce et al. "All-Optical Synthesis of an Arbitrary Linear Transformation Using Diffractive Surfaces", Light Science & Applications, DOI: 10.1038/s41377-021-00623-5. This chapter presents a numerical study that investigates the capabilities of diffractive optical networks in synthesizing arbitrary linear transformations between their input and output fields-of-view.

Spatially-engineered diffractive surfaces have emerged as a powerful framework to control light-matter interactions for e.g., statistical inference and the design of task-specific optical components. Here, we report the design of diffractive surfaces to all-optically perform arbitrary complex-valued linear transformations between an input (N_i) and output (N_o), where N_i and N_o represent the number of pixels at the input and output fields-of-view (FOVs), respectively. First, we consider a single diffractive surface and use a matrix pseudoinverse-based method to determine the complex-valued transmission coefficients of the diffractive features/neurons to all-optically perform a desired/target linear transformation. In addition to this *data-free* design approach, we also consider a deep learning-based design method to optimize the transmission coefficients of diffractive surfaces by using examples of input/output fields corresponding to the target transformation. We compared the all-optical transformation errors and diffraction efficiencies achieved using data-free designs as well as data-driven (deep learning-based) diffractive designs to all-optically perform (*i*) arbitrarily-chosen complex-valued transformations including unitary, nonunitary and noninvertible transforms, (*ii*) 2D discrete Fourier transformation, (*iii*) arbitrary 2D permutation operations, and (*iv*) high-pass filtered coherent

imaging. Our analyses reveal that if the total number (N) of spatially-engineered diffractive features/neurons is $\geq N_i \times N_o$, both design methods succeed in all-optical implementation of the target transformation, achieving negligible error. However, compared to data-free designs, deep learning-based diffractive designs are found to achieve significantly larger diffraction efficiencies for a given N and their all-optical transformations are more accurate for $N < N_i \times N_o$. These conclusions are generally applicable to various optical processors that employ spatially-engineered diffractive surfaces.

5.1 Introduction

It is well-known that optical waves can be utilized for the processing of spatial and/or temporal information[138–143]. Using optical waves to process information is appealing since computation can be performed at the speed of light, with high parallelization and throughput, also providing potential power advantages. For this broad goal, various optical computing architectures have been demonstrated in the literature[144–158]. With the recent advances in photonic material engineering, e.g., metamaterials, metasurfaces and plasmonics, the utilization of advanced diffractive materials that can precisely shape optical wavefronts through light-matter interaction has become feasible[159–164]. For example, optical processors formed through spatially-engineered diffractive surfaces have been shown to achieve both statistical inference and deterministic tasks, such as image classification, single-pixel machine vision and spatially-controlled wavelength division multiplexing, among others[165–174].

Since scalar optical wave propagation in free space and light transmission through diffractive surfaces constitute linear phenomena, the light transmission from an input field-of-view (FOV) to an output FOV that is engineered through diffractive surfaces can be formulated using linear algebra[172]. As a result, together with the free space diffraction, the light transmittance patterns of diffractive surfaces (forming an optical network) collectively define a certain complex-valued all-optical linear transformation between the input and output FOVs. In this paper, we focus on designing these spatial patterns and diffractive surfaces that can all-optically compute a desired, target transformation. We demonstrate that an arbitrary complex-valued linear transformation between an input and output FOV can be realized using spatially-engineered diffractive surfaces, where each feature (neuron) of a diffractive layer modulates the amplitude and/or phase of the optical wave field. In generating the needed diffractive surfaces to all-optically achieve a given

target transformation, we use both a matrix pseudoinverse-based design that is *data-free* as well as a data-driven, deep learning-based design method. In our analysis, we compared the approximation capabilities of diffractive surfaces for performing various all-optical linear transformations as a function of the total number of diffractive neurons, number of diffractive layers and the area of the input/output FOVs. For these comparisons, we used as our target transformations arbitrarily generated complex-valued unitary, nonunitary and noninvertible transforms, 2D Fourier transform, 2D random permutation operation as well as high-pass filtered coherent imaging operations.

Our results reveal that when the total number of engineered/optimized diffractive neurons of a material design exceeds $N_i \times N_o$, both the data-free and data-driven diffractive designs successfully approximate the target linear transformation with negligible error; here, N_i and N_o refer to the number of diffraction-limited, independent spots/pixels located within the area of the input and output FOVs, respectively. This means, to all-optically perform an arbitrary complex-valued linear transformation between larger input and/or larger output FOVs, larger area diffractive layers with more neurons or a larger number of diffractive layers need to be utilized. Our analyses further reveal that deep learning-based data driven diffractive designs (that learn a target linear transformation through examples of input/output fields) overall achieve much better diffraction efficiency at the output FOV. All in all, our analysis confirms that for a given diffractive layer size, with a certain number of diffractive features per layer (like a building block of a diffractive network), the creation of deeper diffractive networks with one layer following another, can improve both the transformation error and the diffraction efficiency of the resulting all-optical transformation.

Our results and conclusions can be broadly applied to any part of the electromagnetic spectrum to design all-optical processors using spatially-engineered diffractive surfaces to perform an arbitrary complex-valued linear transformation.

5.2 Results

Formulation of all-optical transformations using diffractive surfaces

Let $\boldsymbol{i}$ and $\boldsymbol{o}$ be the column vectors that include the samples of the 2D complex-valued input and output FOVs, respectively, as shown in Fig. 5.1.a. Here we assume that the optical wave field can be represented using the scalar field formulation[175–177]. $\boldsymbol{i}$ and $\boldsymbol{o}$ are generated by, first, sampling the 2D input and output FOVs, and then vectorizing the resulting 2D matrices in a column-major order. Following our earlier notation, N_i and N_o represent the number of diffraction-limited spots/pixels on the input and output FOVs, respectively, which also define the lengths of the vectors $\boldsymbol{i}$ and $\boldsymbol{o}$**.** In our simulations, we assume that the sampling period along both the horizontal and vertical directions is $\lambda/2$, where λ is the wavelength of the monochromatic scalar optical field. With this selection in our model, we include all the propagating modes that are transmitted through the diffractive layer(s).

To implement the wave propagation between parallel planes in free space, we generate a matrix, $\boldsymbol{H}_d$, where d is the axial distance between two planes (e.g., $d \geq \lambda$). Since this matrix represents a convolution operation where the 2D impulse response originates from the Rayleigh-Sommerfeld diffraction formulation[140], it is a Toeplitz matrix[178]. We generate this matrix using the Fourier relation in the discrete domain as

$$\boldsymbol{H}_d = \boldsymbol{W}^{-1}\boldsymbol{D}\boldsymbol{W} = \boldsymbol{W}^{H}\boldsymbol{D}\boldsymbol{W} \tag{5.1}$$

where $\boldsymbol{W}$ and $\boldsymbol{W}^{-1}$ are the 2D discrete Fourier transform (DFT) and inverse discrete Fourier transform (IDFT) matrices, respectively, and the superscript *H* represents the matrix Hermitian operation. We choose the scaling constant appropriately such that the unitarity of the DFT operation is preserved, i.e., $\boldsymbol{W}^{-1} = \boldsymbol{W}^{H}$ [178]. The matrix, $\boldsymbol{D}$, represents the transfer function of free space propagation in the 2D Fourier domain and it includes nonzero elements only along its main diagonal entries. These entries are the samples of the function, $e^{jk_z d}$, for $0 \leq k_z = \sqrt{k^2 - \left(k_x^2 + k_y^2\right)} \leq k$, where $k_x, k_y \in [-k, k)$. Here $k = 2\pi/\lambda$ is the wavenumber of the monochromatic optical field and $\left(k_x, k_y\right)$ pair represents the 2D spatial frequency variables along x and y directions, respectively[140]. To ignore the evanescent modes, we choose the diagonal entries of $\boldsymbol{D}$ that correspond to the k_z values for $k^2 \leq k_x^2 + k_y^2$ as zero; since $d \geq \lambda$ this is an appropriate selection. In our model, we choose the 2D discrete wave propagation square window size, $\sqrt{N_d} \times \sqrt{N_d}$, large enough (e.g., $N_d = 144^2$) such that the physical wave propagation between the input plane, diffractive layers and the output plane is simulated accurately [179]. Also, since $\boldsymbol{H}_d$ represents a convolution in 2D space, the entries of $\boldsymbol{W}$, $\boldsymbol{W}^{-1}$ and $\boldsymbol{D}$ follow the same vectorization procedure applied to the input and output FOVs. As a result, the sizes of all these matrices become $N_d \times N_d$.

Since the diffractive surfaces are modeled as thin elements, the light transmission through surfaces can be formulated as a pointwise multiplication operation, where the output optical field of a layer equals to its input optical field multiplied by the complex-valued transmission function, $t(x, y)$, of the diffractive layer. Hence, in matrix formulation, this is represented by a diagonal

matrix $\boldsymbol{T}$, where the diagonal entries are the vectorized samples of $t(x, y)$. Hence the size of $\boldsymbol{T}$ becomes $N_L \times N_L$, where N_L is the total number of diffractive features (referred to as neurons) on the corresponding layer.

We also assume that the forward propagating optical fields are zero outside of the input FOV and outside of the transmissive parts of each diffractive surface, so that we solely analyze the modes that are propagating through the transmissive diffractive layers. This is not a restrictive assumption as it can be simply satisfied by introducing light blocking, opaque materials around the input FOV and the diffractive layers. Furthermore, although the wave field is not zero outside of the output FOV, we only formulate the optical wave propagation between the input and output FOVs since the complex-valued transformation ($\boldsymbol{A}$) that we would like to approximate is *defined* between $\boldsymbol{i}$ and $\boldsymbol{o}$. As a result of these, we delete the appropriate rows and columns of $\boldsymbol{H}_d$, which are generated based on Equation 5.1. We denote the resulting matrix as $\boldsymbol{H}'_d$.

Based on these definitions, the relationship between the input and output FOVs for a diffractive network that has K diffractive layers can be written as

$$\boldsymbol{o}' = \boldsymbol{H}'_{d_{K+1}} \boldsymbol{T}_K \boldsymbol{H}'_{d_K} \cdots \boldsymbol{T}_2 \boldsymbol{H}'_{d_2} \boldsymbol{T}_1 \boldsymbol{H}'_{d_1} \boldsymbol{i} = \boldsymbol{A}' \boldsymbol{i} \tag{5.2}$$

as shown in Fig. 5.1.a. Here d_1 is the axial distance between the input FOV and the first diffractive layer, d_{K+1} is the axial distance between the K^{th} layer and the output FOV, and d_l for $l \in \{2,3,\cdots,K\}$ is the axial distance between the $(l-1)^{th}$ and l^{th} diffractive layers (see Fig. 5.1.a). Also, $\boldsymbol{T}_l$ for $l \in \{1,2,\cdots,K\}$ is the complex-valued light transmission matrix of the l^{th} layer. The size of $\boldsymbol{H}'_{d_1}$ is $N_{L_1} \times N_i$, the size of $\boldsymbol{H}'_{d_{K+1}}$ is $N_o \times N_{L_K}$ and the size of $\boldsymbol{H}'_{d_l}$ is $N_{L_l} \times N_{L_{l-1}}$ for $l \in \{2,3,\cdots,K\}$, where N_{L_l} is the number of diffractive neurons at the l^{th} diffractive

layer. Note that, in our notation in Equation 5.2, we define $\boldsymbol{o}'$ as the calculated output by the diffractive system, whereas $\boldsymbol{o}$ refers to the ground truth/target output in response to $\boldsymbol{i}$. The matrix $\boldsymbol{A}'$ in Equation 5.2, that is formed by successive diffractive layers/surfaces, represents the all-optical transformation performed by the diffractive network from the input FOV to the output FOV. Note that this formalism does not aim to optimize the diffractive system in order to implement only one given pair of input-output complex fields; instead it aims to all-optically approximate an arbitrary complex-valued linear transformation, $\boldsymbol{A}$.

Matrix pseudoinverse-based synthesis of an arbitrary complex-valued linear transformation using a single diffractive surface ($\boldsymbol{K = 1}$)

In this section, we focus on data-free design of a single diffractive layer ($K = 1$), in order to determine the diagonal entries of $\boldsymbol{T}_1$ such that the resulting transformation matrix, $\boldsymbol{A}'$ which is given by Equation 5.2, approximates the transformation matrix $\boldsymbol{A}$. To accomplish this, we first vectorize $\boldsymbol{A}'$ in a column-major order and write it as[172]

$$\begin{aligned} vec(\boldsymbol{A}') = \boldsymbol{a}' &= vec\left(\boldsymbol{H}'_{d_2}\boldsymbol{T}_1\boldsymbol{H}'_{d_1}\right) \\ &= \left(\boldsymbol{H}'^{T}_{d_1} \otimes \boldsymbol{H}'_{d_2}\right) vec(\boldsymbol{T}_1) \end{aligned} \tag{5.3}$$

where $\otimes$ and the superscript T represent the Kronecker product and the matrix transpose operator, respectively. Since the elements of $vec(\boldsymbol{T}_1)$ are nonzero only for the diagonal elements of $\boldsymbol{T}_1$, Equation 5.3 can be further simplified as

$$\boldsymbol{a}' = \boldsymbol{H}'\boldsymbol{t}_1 \tag{5.4}$$

where $\boldsymbol{t}_1[l] = \boldsymbol{T}_1[l,l]$ and $\boldsymbol{H}'[:,l] = \boldsymbol{H}'^{T}_{d_1}[:,l] \otimes \boldsymbol{H}'_{d_2}[:,l]$ for $l \in \{1,2, \cdots, N_{L_1}\}$, and $[:,l]$ represents the l^{th} column of the associated matrix in our notation. Here the matrix $\boldsymbol{H}'$ has size $N_i N_o \times N_{L_1}$ and is a full-rank matrix with rank $D = min(N_i N_o, N_{L_1})$ for $d_1 \neq d_2$. If $d_1 = d_2$, the maximum rank reduces to $N_i(N_i + 1)/2$ when $N_i = N_o$ [172]. We assume that $d_1 \neq d_2$ and denote the maximum achievable rank as D_{max}, which is equal to $N_i N_o$.

Based on Equation 5.4, the computation of the neuron transmission values of the diffractive layer that approximates a given complex-valued transformation matrix $\boldsymbol{A}$ can be reduced to an L2-norm minimization problem, where the approximation error which is subject to the minimization is [178]

$$\begin{aligned} \|\boldsymbol{a} - m\boldsymbol{a}'\|^2 = \|\boldsymbol{a} - m\boldsymbol{H}'\boldsymbol{t}_1\|^2 = \|\boldsymbol{a} - \boldsymbol{H}'\hat{\boldsymbol{t}}_1\|^2 = \|\boldsymbol{a} - \hat{\boldsymbol{a}}'\|^2 \\ = \frac{1}{N_i N_o} \sum_{l=1}^{N_i N_o} |\boldsymbol{a}[l] - m\boldsymbol{a}'[l]|^2 \\ = \frac{1}{N_i N_o} \sum_{l=1}^{N_i N_o} |\boldsymbol{a}[l] - \hat{\boldsymbol{a}}'[l]|^2 \end{aligned} \tag{5.5}$$

where $\boldsymbol{a}$ is the vectorized form of the target transformation matrix $\boldsymbol{A}$, i.e., $vec(\boldsymbol{A}) = \boldsymbol{a}$. We included a scalar, normalization coefficient (m) in Equation 5.5 so that the resulting difference term does not get affected by a diffraction-efficiency related scaling mismatch between $\boldsymbol{A}$ and $\boldsymbol{A}'$; also note that we assume a passive diffractive layer without any optical gain, i.e., $|\boldsymbol{t}_1[l]| \leq 1$ for all $l \in \{1,2, \cdots, N_{L_1}\}$. As a result of this, we also introduced in Equation 5.5, $m\boldsymbol{t}_1 = \hat{\boldsymbol{t}}_1$.

Throughout the paper, we use $\left\|\boldsymbol{A}-\widehat{\boldsymbol{A}}'\right\|^2$ and $\|\boldsymbol{a}-\widehat{\boldsymbol{a}}'\|^2$ interchangeably both referring to Equation 5.5 and define them as the ***all-optical transformation error***. We also refer to $\boldsymbol{a}$, $\boldsymbol{a}'$ and $\widehat{\boldsymbol{a}}'$ as the target transformation (ground truth), estimate transformation and normalized estimate transformation vectors, respectively.

If $N_{L_1} > N_i N_o$, the number of equations in Equation 5.4 becomes less than the number of unknowns and the matrix-vector equation corresponds to an underdetermined system. If $N_{L_1} < N_i N_o$, on the other hand, the equation system becomes an overdetermined system. In the critical case, where $N_{L_1} = N_i N_o$, $\boldsymbol{H}'$ becomes a full-rank square matrix, hence, is an invertible matrix. There are various numerical methods for solving the formulated matrix-vector equation and minimizing the transformation error given in Equation 5.5 [178]. In this paper, we adopt the pseudoinverse-based method among other numerical methods in computing the neuron transmission values in finding the estimate transformation $\boldsymbol{A}'$ for all the cases, i.e., $N_{L_1} > N_i N_o$, $N_{L_1} < N_i N_o$ and $N_{L_1} = N_i N_o$. For this, we compute the neuron values from a given target transformation as

$$\hat{\boldsymbol{t}}_1 = \boldsymbol{H}'^{\dagger}\boldsymbol{a} \tag{5.6}$$

where $\boldsymbol{H}'^{\dagger}$ is the pseudoinverse of $\boldsymbol{H}'$. This pseudoinverse operation is performed using the singular value decomposition (SVD) as

$$\boldsymbol{H}'^{\dagger} = \boldsymbol{V}\boldsymbol{\Sigma}^{-1}\boldsymbol{U}^{H} \tag{5.7}$$

where $\boldsymbol{U}$ and $\boldsymbol{V}$ are orthonormal matrices and $\boldsymbol{\Sigma}$ is a diagonal matrix that contains the singular values of $\boldsymbol{H}'$. $\boldsymbol{U}, \boldsymbol{V}$ and $\boldsymbol{\Sigma}$ form the $\boldsymbol{H}'$ matrix as

$$\boldsymbol{H}' = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^H \tag{5.8}$$

To prevent the occurrence of excessively large numbers that might cause numerical artifacts, we take very small singular values as zero during the computation of $\boldsymbol{\Sigma}^{-1}$. After computing $\hat{\boldsymbol{t}}_1$, the normalization constant (m) and the physically realizable neuron values can be calculated as:

$$m = \max_l(|\hat{\boldsymbol{t}}_1|) \quad \text{and} \quad \boldsymbol{t}_1 = \hat{\boldsymbol{t}}_1/m \tag{5.9}$$

In summary, the vector $\boldsymbol{t}_1$ that includes the transmittance values of the diffractive layer is computed from a given, target transformation vector, $\boldsymbol{a}$, using Equations 5.6 and 5.9, and then the resulting estimate transformation vector, $\boldsymbol{a}'$, is computed using Equation 5.4. Finally, $\boldsymbol{A}'$, is obtained from $\boldsymbol{a}'$ by reversing the vectorization operation.

Deep learning-based synthesis of an arbitrary complex-valued linear transformation using diffractive surfaces ($\boldsymbol{K \geq 1}$)

Different from the numerical pseudoinverse-based design method described in the previous section, which is data-free in its computational steps, deep learning-based design of diffractive layers utilize a training dataset containing examples of input/output fields corresponding to a target transformation $\boldsymbol{A}$. In a K-layered diffractive network, our optical forward model implements Equation 5.2, where the diagonal entries of each $\boldsymbol{T}_l$ matrix for $l \in \{1,2, \cdots, K\}$ become the arguments subject to the optimization. At each iteration of deep

learning-based optimization during the error-back-propagation algorithm, the complex-valued neuron values are updated to minimize the following normalized mean-squared-error loss function:

$$\left\|\widetilde{\boldsymbol{o}}_s - \widetilde{\boldsymbol{o}}'_{s,c}\right\|^2 = \frac{1}{N_o}\sum_{l=1}^{N_o}\left|\sigma_s \boldsymbol{o}_s[l] - \sigma'_{s,c}\boldsymbol{o}'_{s,c}[l]\right|^2 \tag{5.10}$$

where

$$\boldsymbol{o}_s = \boldsymbol{A}\boldsymbol{i}_s \quad \text{and} \quad \boldsymbol{o}'_{s,c} = \boldsymbol{A}'_c\boldsymbol{i}_s \tag{5.11}$$

refer to the ground truth and the estimated output field by the diffractive network, respectively, for the s^{th} input field in the training dataset, $\boldsymbol{i}_s$. The subscript c indicates the current state of the all-optical transformation at a given iteration of the training that is determined by the current transmittance values of the diffractive layers. The constant σ_s normalizes the energy of the ground truth field at the output FOV and can be written as

$$\sigma_s = \left(\sum_{l=1}^{N_o}|\boldsymbol{o}_s[l]|^2\right)^{-\frac{1}{2}} \tag{5.12}$$

Also, the complex valued $\sigma'_{s,c}$ is calculated to minimize Equation 5.10. It can be computed by taking the derivative of $\left\|\widetilde{\boldsymbol{o}}_s - \widetilde{\boldsymbol{o}}'_{s,c}\right\|^2$ with respect to $\sigma'^{*}_{s,c}$, which is the complex conjugate of $\sigma'_{s,c}$, and then equating the resulting expression to zero,[180] which yields:

$$\sigma'_{s,c} = \frac{\sum_{l=1}^{N_o} \sigma_s \boldsymbol{o}_s[l] \boldsymbol{o}'^{*}_{s,c}[l]}{\sum_{l=1}^{N_o} \left|\boldsymbol{o}'_{s,c}[l]\right|^2} \tag{5.13}$$

After the training is over, which is a one-time effort, the estimate transformation matrix and the corresponding vectorized form, $\boldsymbol{A}'$ and $vec(\boldsymbol{A}') = \boldsymbol{a}'$, are computed using the optimized neuron transmission values in Equation 5.2. After computing $\boldsymbol{a}'$, we also compute the normalization constant, m, which minimizes $\|\boldsymbol{a} - m\boldsymbol{a}'\|^2$, resulting in:

$$m = \frac{\sum_{l=1}^{N_i N_o} \boldsymbol{a}[l] \boldsymbol{a}'^{*}[l]}{\sum_{l=1}^{N_i N_o} |\boldsymbol{a}'[l]|^2} \tag{5.14}$$

In summary, an optical network that includes K diffractive surfaces can be optimized using deep learning through training examples of input/output fields that correspond to a target transformation, $\boldsymbol{A}$. Starting with the next section, we will analyze and compare the resulting all-optical transformations that can be achieved using data-driven (deep learning-based) as well as data-free designs that we introduced.

Comparison of all-optical transformations performed through diffractive surfaces designed by matrix pseudoinversion vs. deep learning

In this section we present a quantitative comparison of the pseudoinverse- and deep learning-based methods in synthesizing various all-optical linear transformations between the input and output FOVs using diffractive surfaces. In our analysis, we took the total number of pixels in both the input and output FOVs as $N_i = N_o = 64$ ($i.e.\ 8 \times 8$) and the size of each $\boldsymbol{H}_d$ matrix was $144^2 \times 144^2$ with $N_d = 144^2$. The linear transformations that we used as our comparison testbeds are (*i*) arbitrarily generated complex-valued unitary transforms, (*ii*)

arbitrarily generated complex-valued nonunitary and invertible transforms, (*iii*) arbitrarily generated complex-valued noninvertible transforms, (*iv*) the 2D discrete Fourier transform, (*v*) a permutation matrix-based transformation, and (*vi*) a high-pass filtered coherent imaging operation. The details of the diffractive network configurations, training image datasets, training parameters, computation of error metrics and the generation of ground truth transformation matrices are presented in Section 5.4. Next, we present the performance comparisons for different all-optical transformations.

Case 1: Arbitrary complex-valued unitary and nonunitary transforms

In Figs. 5.1-5.3, we present the results for an arbitrarily selected complex-valued *unitary* transforms that is approximated using diffractive surface designs with different number of diffractive layers, K, and different number of neurons, $N = \sum_{k=1}^{K} N_{L_k}$. Similarly, Figs. 5.4-5.6 report a different *nonunitary*, arbitrarily selected complex-valued linear transforms performed through diffractive surface designs. To cover different types of transformations, Figs. 5.4-5.6 report an invertible nonunitary and a noninvertible (hence, nonunitary) transformation, respectively. The magnitude and phase values of these target transformations ($\boldsymbol{A}$) are also shown in Figs. 5.1.b, 5.b.

To compare the performance of all-optical transformations that can be achieved by different diffractive designs, Fig. 5.1.c and Fig. 5.4.c report the resulting transformation errors for the above described testbeds ($\boldsymbol{A}$ matrices) as a function of N and K. It can be seen that, in all of the diffractive designs reported in these figures, there is a monotonic decrease in the transformation error as the total number of neurons in the network increases. In data-free, matrix pseudoinverse-based designs ($K = 1$), the transformation error curves reach a baseline,

approaching ~0 starting at $N = 64^2$. This empirically-found turning point of the transformation error at $N = 64^2$ also coincides with the limit of the information processing capacity of the diffractive network dictated by $D_{max} = N_i N_o = 64^2$ [172]. Beyond this point, i.e., for $N > 64^2$, the all-optical transformation errors of data-free diffractive designs remain negligible for these complex-valued unitary as well as nonunitary transformations defined in Fig. 5.1.b and Fig. 5.4.b.

On the other hand, for data-driven, deep learning-based diffractive designs, one of the key observations is that, as the number of diffractive layers (K), increases, the all-optical transformation error decreases for the same N. Stated differently, deep learning-based, data-driven diffractive designs prefer to distribute/divide the total number of neurons (N) into different, successive layers as opposed having all the N neurons at a single, large diffractive layer; the latter, deep learning-designed $K = 1$, exhibits much worse all-optical transformation error compared to e.g., $K = 4$ diffractive layers despite the fact that both of these designs have the same number of trainable neurons (N). Furthermore, as illustrated in Fig. 5.1 and Fig. 5.4, deep learning-based diffractive designs with $K = 4$ layers match the transformation error performance of data-free designs based on matrix pseudoinversion and also exhibit negligible transformation error for $N \geq 64^2 = N_i N_o$. However, when $N < 64^2$ the deep learning-based diffractive designs with $K = 4$ layers achieve smaller transformation errors compared to data-free diffractive designs that have the same number of neurons. Similar conclusions can be made in Figs. 5.1.e and 5.4.e, by comparing the mean-squared-error (MSE) values calculated at the output FOV using test images (input fields). For $N \geq 64^2 = N_i N_o$ the deep learning-based diffractive designs ($K = 4$) along with the data-free diffractive designs achieve output MSE values that approach ~0, similar to the all-optical transformation errors that approach ~0 in Fig.

5.1.c and Fig. 5.4.c. However for designs that have smaller number of neurons, i.e., $N < N_i N_o$, the deep learning-based diffractive designs with $K = 4$ achieve much better MSE at the output FOV compared to data-free diffractive designs that have same number of neurons (N).

In addition to these, one of the most significant differences between the pseudoinverse-based data-free diffractive designs and deep learning-based counterparts is observed in the optical diffraction efficiencies calculated at the output FOV; see Figs. 5.1.f and 5.4.f. Even though the transformation errors (or the output MSE values) of the two design approaches remain the same (~0) for $N \geq 64^2 = N_i N_o$, the diffraction efficiencies of the all-optical transformations learned using deep learning significantly outperform the diffraction efficiencies achieved using data-free, matrix pseudoinverse-based designs as shown in Figs. 5.1.f and 5.4.f.

On top of transformation error, output MSE and diffraction efficiency metrics, Figs. 5.1.d, and 5.4.d also report the cosine similarity (see Section 4) between the estimated all-optical transforms and the ground truth (target) transforms. These cosine similarity curves show the same trend and support the same conclusions as with the transformation error curves reported earlier; this is not surprising as the transformation error and cosine similarity metrics are analytically related to each other as detailed later. For $N \geq 64^2 = N_i N_o$, the cosine similarity approaches 1, matching the target transformations using both the data-free ($K = 1$) and deep learning-based ($K = 4$) diffractive designs as shown in Figs. 5.1.d and 5.4.d.

To further shed light on the performance of these different diffractive designs, the estimated transformations and their differences (in phase and amplitude) from the target matrices ($\boldsymbol{A}$) are shown in Figs. 5.2 and 5.5 for different diffractive parameters. Similarly, examples of complex-valued input-output fields for different diffractive designs are compared in Figs. 5.3 and

5.6 against the ground truth output fields (calculated using the target transformations), along with the resulting phase and amplitude errors at the output FOV. From these figures, it can be seen that both data-free ($K = 1$) and deep learning-based ($K = 4$) diffractive designs with the same total number of neurons can all-optically generate the desired transformation and output field patterns with negligible error when $N \geq 64^2 = N_i N_o$. For $N < N_i N_o$, on the other hand, the output field amplitude and phase profiles using deep learning-based diffractive designs show much better match to the ground truth output field profiles when compared to data-free, matrix pseudoinverse-based diffractive designs (see e.g., Figs. 5.3, 5.6).

In this subsection, we presented diffractive designs that successfully approximated arbitrary complex-valued transformations, where the individual elements of target $\boldsymbol{A}$ matrices (shown in Fig. 5.1.b and Fig. 5.4.b) were randomly and independently generated as described in Section 4.4. Our results confirm that, for a given total number of diffractive features/neurons (N) available, building deeper diffractive networks where these neurons are distributed across multiple, successive layers, one following the other, can significantly improve the transformation error, output field accuracy and the diffraction efficiency of the whole system to all-optically implement an arbitrary, complex-valued target transformation between an input and output FOV. Starting with the following subsection, we focus on some task-specific all-optical transformations, which are frequently used in various optics and photonics applications.

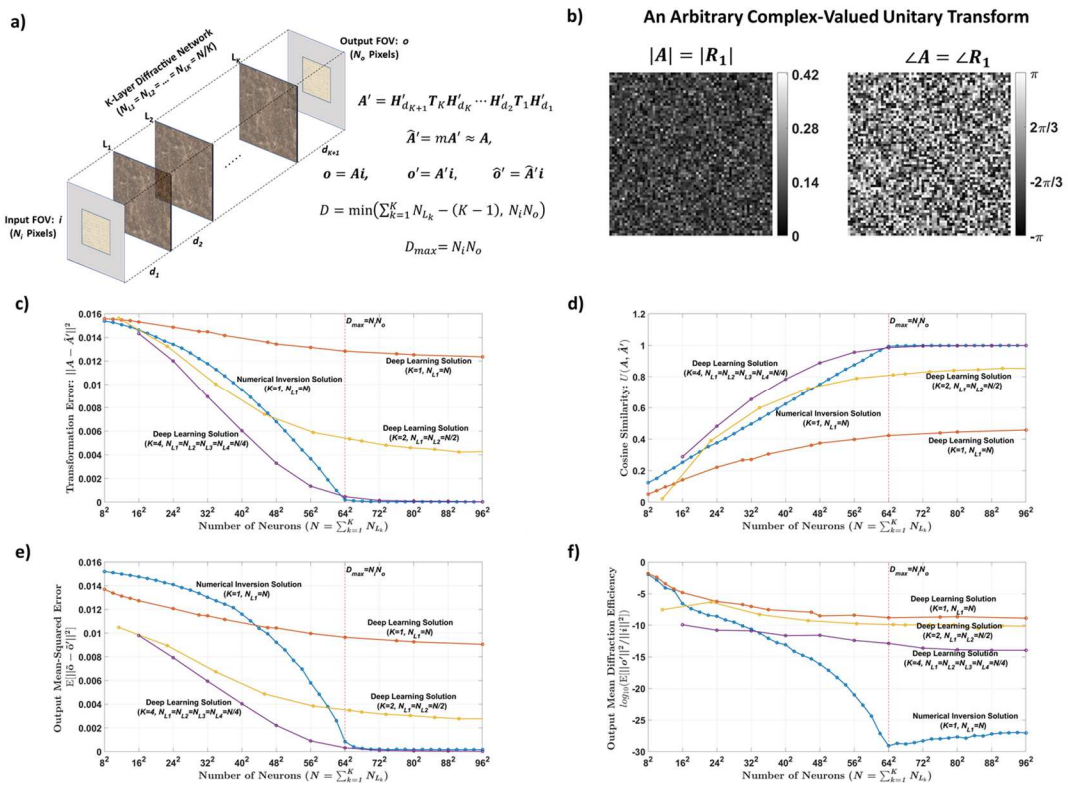

Fig. 5.1 Diffractive all-optical transformation results for an arbitrary complex-valued unitary transform. a. Schematic of a K-layer diffractive network, that all-optically performs a linear transformation between the input and output fields-of-views that have N_i and N_o pixels, respectively. The all-optical transformation matrix due to the diffractive layer(s) is given by A'. b. The magnitude and phase of the ground truth (target) input-output transformation matrix, which is an arbitrarily generated complex-valued unitary transform, i.e., $R_1^H R_1 = R_1 R_1^H = I$. c. All-optical transformation errors (see Equation 5.5). The x-axis of the figure shows the total number of neurons (N) in a K-layered diffractive network, where each diffractive layer includes N/K neurons. Therefore, for each point on the x-axis, the comparison among different diffractive designs (colored curves) is fair as each diffractive design has the same total number of neurons available. The simulation data points are shown with dots and the space between the dots are linearly interpolated. d. Cosine similarity between the vectorized form of the target transformation matrix in (b) and the resulting all-optical transforms (see Equation 5.16). e. Output MSE between the ground-truth output fields and the estimated output fields by the diffractive network (see Equation 5.18). f. The diffraction efficiency of the designed diffractive networks (see Equation 5.19).

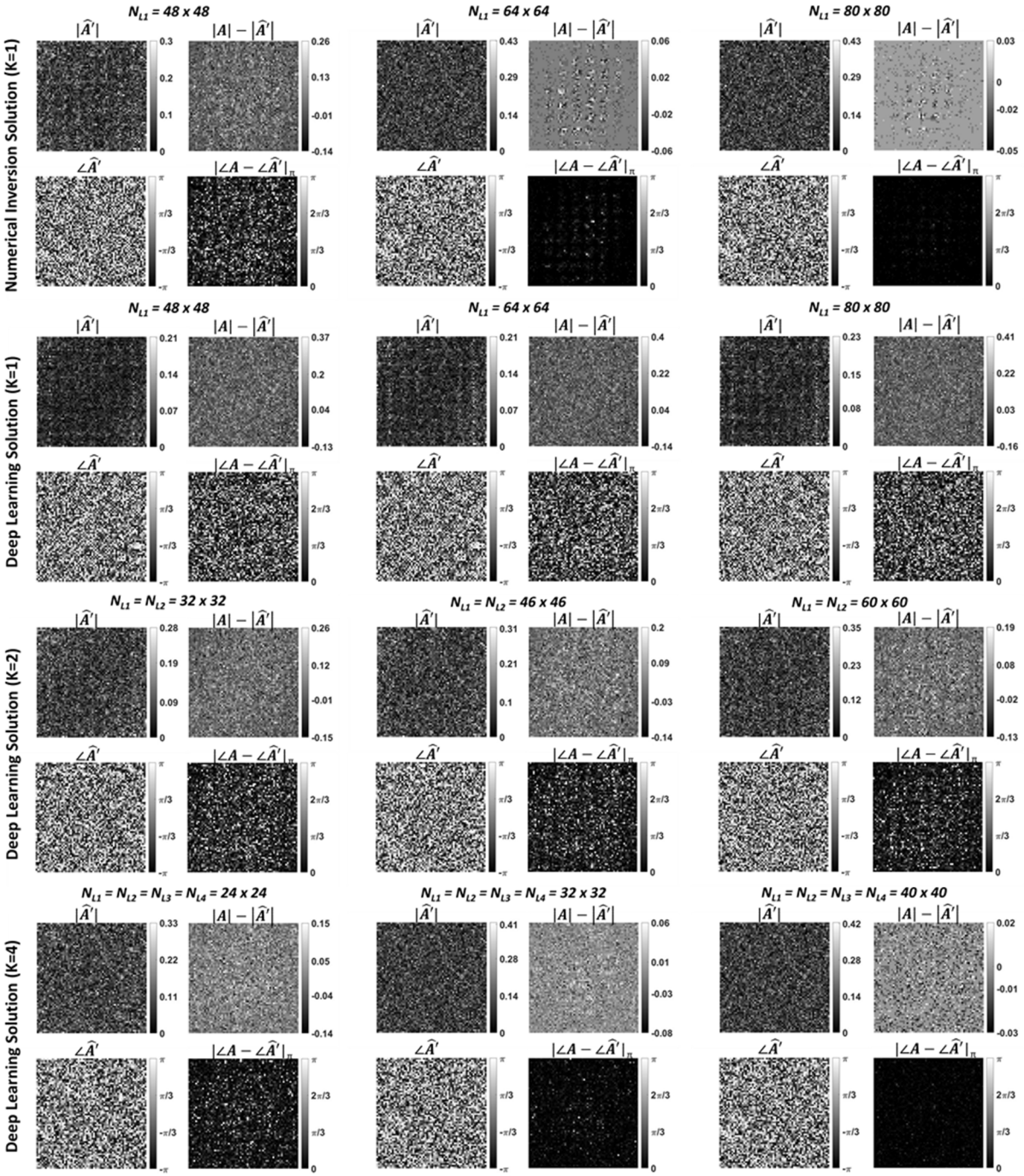

Fig. 5.2 Diffractive all-optical transformations and their differences from the ground truth, target transformation ($\boldsymbol{A}$) presented in Fig. 5.1.b. $\left|\angle A - \angle\hat{A}'\right|_\pi$ indicates the wrapped phase difference between the ground truth and the normalized all-optical transformation.

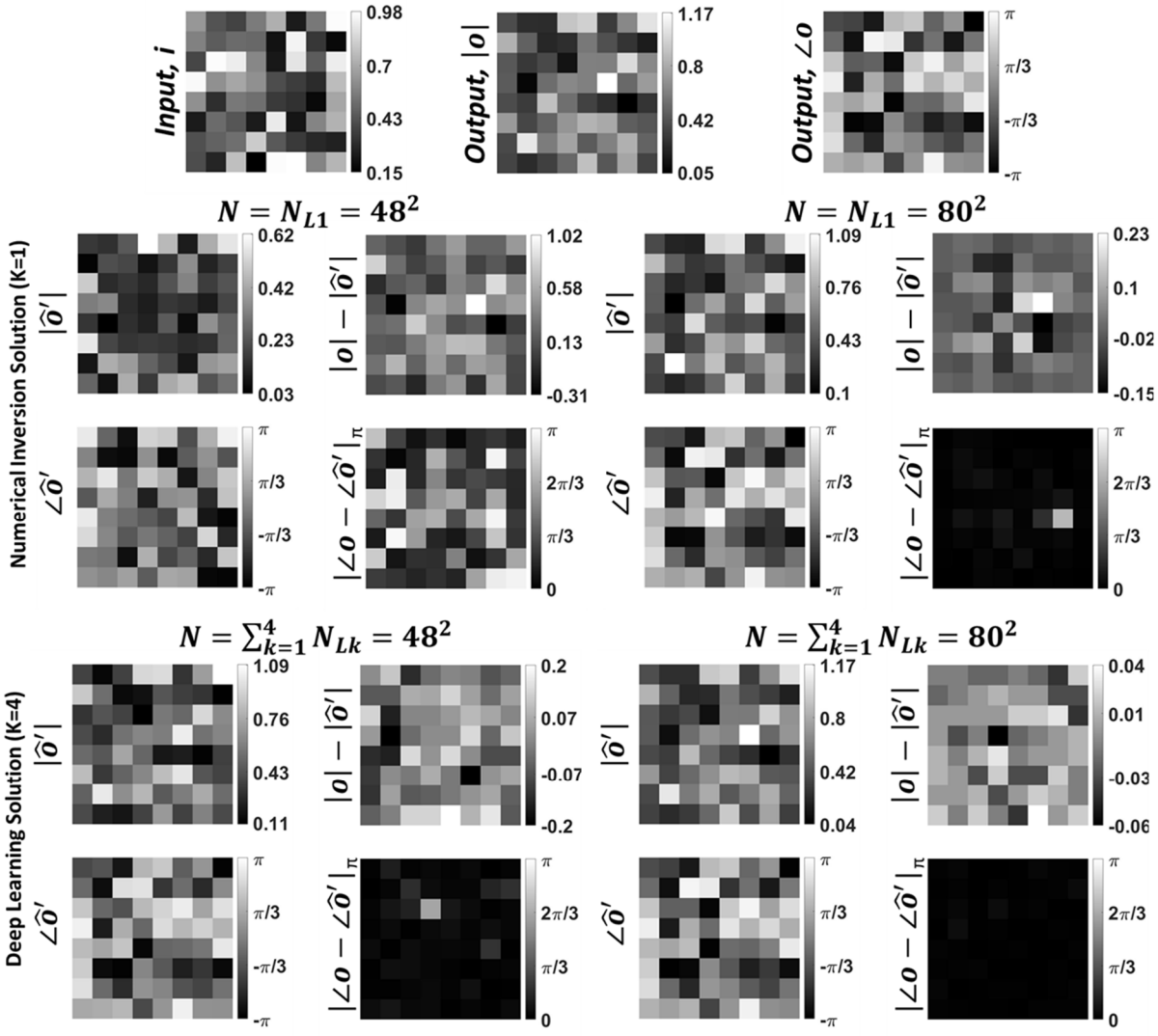

Fig. 5.3 Sample input-output images for the ground truth transformation presented in Fig. 5.1.b and the optical outputs by the diffractive designs for two different choices of N ($N = 48^2$ and $N = 80^2$). The magnitude and phase of the normalized output fields and the differences of these quantities with respect to the ground truth are shown. $|\angle o - \angle\hat{o}'|_\pi$ indicates wrapped phase difference between the ground truth and the normalized output field.

Case 2: 2D discrete Fourier transform

Here we show that the 2D Fourier transform operation can be performed using diffractive designs such that the complex field at output FOV reflects the 2D discrete Fourier transform of the input field. Compared to lens-based standard Fourier transform operations, diffractive surface-based solutions are not based on the paraxial approximation and offer a much more compact set-up (with a significantly smaller axial distance, e.g., $< 50\lambda$, between the input-output planes) and do not suffer from aberrations, which is especially important for larger input/output FOVs.

The associated transform matrix ($\boldsymbol{A}$) corresponding to 2D discrete Fourier transform, all-optical transformation error, cosine similarity of the resulting all-optical transforms with respect the ground truth, the output MSE and the diffraction efficiency are shown in Fig. 5.7. For all these curves and metrics, our earlier conclusions made in Section 2.4.1 are also applicable. Data-free ($K = 1$) and deep learning-based ($K = 4$) diffractive designs achieve accurate results at the output FOV for $N \geq N_i N_o = 64^2$, where the transformation error and the output MSE both approach to $\sim$0 while the cosine similarity reaches $\sim$1, as desired. In terms of the diffraction efficiency at the output FOV, similar to our earlier observations in the previous section, deep learning-based diffractive designs offer major advantages over data-free diffractive designs. Further advantages of deep learning-based diffractive designs over their data-free counterparts include significantly improved output MSE and reduced transformation error for $N < N_i N_o$, confirming our earlier conclusions made in Section 2.4.1.

To further show the success of the diffractive designs in approximating the 2D discrete Fourier transformation, in Fig. 5.8 we report the estimated transformations and their differences (in phase and amplitude) from the target 2D discrete Fourier transformation matrix for different diffractive designs. Furthermore, in Fig. 5.9, examples of complex-valued input-output fields for different diffractive designs are compared against the ground truth output fields (calculated using the 2D discrete Fourier transformation), along with the resulting phase and amplitude errors at the output FOV, all of which illustrate the success of the presented diffractive designs.

Case 3: Permutation matrix

For a given randomly generated permutation matrix ($\boldsymbol{P}$), the task of the diffractive design is to all-optically obtain the permuted version of the input complex-field at the output FOV. Although the target ground truth matrix ($\boldsymbol{P}$) for this case is real-valued and relatively simpler compared to that of e.g., the 2D Fourier transform matrix, an all-optical permutation operation that preserves the phase and amplitude of each point is still rather unconventional and challenging to realize using standard optical components. To demonstrate this capability, we randomly selected a permutation matrix as shown in Fig. 5.10b and designed various diffractive surfaces to all-optically perform this target permutation operation at the output FOV. The performances of these data-free and data-driven, deep learning-based diffractive designs are compared in Figs. 5.10c-f. The success of the diffractive all-optical transformations, matching the target permutation operation is demonstrated when $N \geq N_i N_o$, revealing the same conclusions discussed earlier for the other transformation matrices that were tested. For example, deep learning-based diffractive designs ($K = 4$) with $N \geq N_i N_o$ neurons were successful in performing the randomly selected permutation operation all-optically, and achieved a transformation error and output MSE of ~ 0, together with a cosine similarity of ~ 1 (see Fig.

5.10). Estimated transforms and sample output patterns, together with their differences with respect to the corresponding ground truths are also reported in Figs. 5.11 and 5.12, respectively, further demonstrating the success of the presented diffractive designs.

Case 4: High-pass filtered coherent imaging

In this sub-section, we present diffractive designs that perform high-pass filtered coherent imaging, as shown in Fig. 5.13. This high-pass filtering transformation is based on the Laplacian operator described in Section 5.4. Similar to the 2D discrete Fourier transform operation demonstrated earlier, diffractive surface-based solutions to high-pass filtered coherent imaging are not based on a low numerical aperture assumption or the paraxial approximation, and provide an axially compact implementation with a significantly smaller distance between the input-output planes (e.g., $< 50\lambda$); furthermore, these diffractive designs can handle large input/output FOVs without suffering from aberrations.

Our results reported in Fig. 5.13 also exhibit a similar performance to the previously discussed all-optical transformations, indicating that the pseudoinverse-based diffractive designs and the deep learning-based designs are successful in their all-optical approximation of the target transformation, reaching a transformation error and output MSE of ~0 for $N \geq N_i N_o$. Same as in other transformations that we explored, deep learning-based designs offer significant advantages compared to their data-free counterparts in the diffraction efficiency that is achieved at the output FOV. The estimated sample transformations and their differences from the ground truth transformation are shown in Fig. 5.14. Furthermore, as can be seen from the estimated output images and their differences with respect to the corresponding ground truth images (shown in Fig. 5.15), the diffractive designs can accurately perform high-pass filtered coherent imaging for $N \geq$

$N_i N_o$, and for $N < N_i N_o$ deep learning-based diffractive designs exhibit better accuracy in approximating the target output field, which are in agreement with our former observations in earlier sections.

5.3 Discussion

Through our results and analysis, we showed that it is possible to synthesize an arbitrary complex-valued linear transformation all-optically using diffractive surfaces. We covered a wide range of target transformations, starting from rather general cases, e.g. arbitrarily generated unitary, nonunitary (invertible) and noninvertible transforms, also extending to more specific transformations such as the 2D Fourier transform, 2D permutation operation as well as high-pass filtered coherent imaging operation. In all the linear transformations that we presented in this paper, the diffractive networks realized the desired transforms with negligible error when the total number of neurons reached $N \geq N_i N_o$. It is also important to note that the all-optical transformation accuracy of the deep learning-based diffractive designs improves as the number of diffractive layers is increased, e.g., from $K = 1, 2$ to $K = 4$. Despite sharing the same number of total neurons in each case (i.e., $N = \sum_{k=1}^{K} N_{L_k}$), deep learning-based diffractive designs prefer to distribute these N trainable diffractive features/neurons into multiple layers, favoring deeper diffractive designs overall.

In addition to the all-optical transformation error, cosine similarity and output MSE metrics, the output diffraction efficiency is another very important metric as it determines the signal-to-noise ratio of the resulting all-optical transformation. When we compare the diffraction efficiencies of different networks, we observe that the data-free, matrix pseudoinverse-based designs perform the worst among all the configurations that we have explored (see Figs. 5.1.f,

5.4.f, 5.7.f, 5.10.f, and 5.13.f). This is majorly caused by the larger magnitudes of the transmittance values of the neurons that are located close to the edges of the diffractive layer, when compared to the neurons at the center of the same layer. Since these "edge" neurons are further away from the input FOV, their larger transmission magnitudes ($|\boldsymbol{t}|$) compensate for the significantly weaker optical power that falls onto these edge neurons from the input FOV. Since we are considering here passive diffractive layers only, the magnitude of the transmittance value of an optical neuron cannot be larger than one (i.e., $|\boldsymbol{t}| \leq 1$), and therefore as the edge neurons in a data-free design start to get more transmissive to make up for the weaker input signals at their locations, the transmissivity of the central neurons of the diffractive layer become lower, balancing off their relative powers at the output FOV to be able to perform an arbitrary linear transformation. This is at heart of the poor diffraction efficiency that is observed with data-free, matrix pseudoinverse-based designs. In fact, the same understanding can also intuitively explain why deep learning-based diffractive designs prefer to distribute their trainable diffractive neurons into multiple layers. By dividing their total trainable neuron budget (N) into multiple layers, deeper diffractive designs (e.g., $K = 4$) avoid using neurons that are laterally further away from the center. This way, the synthesis of an arbitrary all-optical transformation can be achieved much more efficiently, without the need to weaken the transmissivity of the central neurons of a given layer. Stated differently, deep learning-based diffractive designs utilize a given neuron budget more effectively and can efficiently perform an arbitrary complex-valued transformation between an input and output FOV.

In fact, deep learning-based, data-driven diffractive designs can be made even more photon efficient by restricting each diffractive layer to be a phase-only element (i.e., $|\boldsymbol{t}| = 1$ for all the neurons) during the iterative learning process of a target complex-valued transformation. To demonstrate this capability with increased diffraction efficiency, we also designed diffractive networks with phase-only layers. These results indicate that much better output diffraction efficiencies can be achieved using phase-only diffractive networks, with some trade-off in the all-optical transformation performance. The relative increase in the transformation errors and the output MSE values that we observed in phase-only diffractive networks is caused by the reduced degrees of freedom in the diffractive design since $|\boldsymbol{t}| = 1$ for all the neurons. Regardless, by increasing the total number of neurons ($N > N_i N_o$), the phase-only diffractive designs approach the all-optical transformation performance of their complex-valued counterparts designed by deep learning, while also providing a much better diffraction efficiency at the output. Note also that, while the phase-only diffractive layers are individually lossless, the forward propagating optical fields still experience some power loss due to the opaque regions that are assumed to surround the diffractive surfaces (which is a design constraint as detailed in Section 2.1). In addition to these losses, the field energy that lies outside of the output FOV is also considered a loss from the perspective of the target transformation, which is defined between the input and output FOVs.

In our analysis reported so far, there are some practical factors that are not taken into account as part of our forward optical model, which might degrade the performance of diffractive networks: (1) material absorption, (2) surface reflections and (3) fabrication imperfections. By using materials with low loss and appropriately selected 3D fabrication methods, these effects can be made negligible compared with the optical power of the forward

propagating modes within the diffractive network. Alternatively, one can also include such absorption- and reflection-related effects as well as mechanical misalignments (or fabrication imperfections) as part of the forward model of the optical system, which can be better taken into account during the deep learning-based optimization of the diffractive layers. Importantly, previous experimental studies[165–174] reported on various diffractive network applications indicate that the impact of fabrication errors, reflection and absorption-based losses are indeed small and do not create a significant discrepancy between the predictions of the numerical forward models and the corresponding experimental measurements.

Finally, we should emphasize that for diffractive networks that have more than one layer, the transmittance values of the neurons of different layers appear in a coupled, multiplicative nature within the corresponding matrix-vector formulation of the all-optical transformation between the input and output FOVs [172]. Hence, a one-step, matrix pseudoinverse-based design strategy cannot be applied for multi-layered diffractive networks in finding all the neuron transmittance values. Moreover, for diffractive designs with a large N, the sizes of the matrices that need to undergo the pseudoinverse operation grow exponentially, which drastically increases the computational load and may prevent performing matrix pseudoinverse computations due to limited computer memory and computation time. This also emphasizes another important advantage of the deep learning-based design methods which can handle much larger number of diffractive neurons to be optimized for a given target transformation, thanks to the efficient error back-propagation algorithms and computational tools that are available. Similarly, if N_i and N_o are increased as the sizes of the input and output FOVs are enlarged, the total number of diffractive neurons needed to successfully approximate a given complex-valued linear transformation will accordingly increase to $D = N_i N_o$, which indicates the critical number of

total diffractive features (marked with the vertical dashed lines in our performance metrics related figures).

5.4 Materials and Methods

Diffractive network configurations

In our numerical simulations, the chosen input and output FOV sizes are both 8×8 pixels. Hence, the target linear transforms, i.e., $\boldsymbol{A}$ matrices, have a size of $N_o \times N_i = 64 \times 64$. For a diffractive design that has a single diffractive surface ($K = 1$), the chosen axial distances are $d_1 = \lambda$ and $d_2 = 4\lambda$. For the networks that have two diffractive surfaces ($K = 2$), the chosen axial distances are $d_1 = \lambda$ and $d_2 = d_3 = 4\lambda$. Finally, for the 4-layered diffractive network ($K = 4$), the axial distances are chosen as $d_1 = \lambda$ and $d_2 = d_3 = d_4 = 4\lambda$. These axial distances can be arbitrarily changed without changing the conclusions of our analysis; they were chosen large enough to neglect the near-field interactions between successive diffractive layers, and small enough to perform optical simulations with a computationally feasible wave propagation window size. We chose our 2D wave propagation window as $N_d \times N_d = 144 \times 144$, which ends up with a size of $144^2 \times 144^2$ for $\boldsymbol{H}_d$ matrices, resulting in ~430 Million entries in each $\boldsymbol{H}_d$ matrix.

Image datasets and diffractive network training parameters

To obtain the diffractive surface patterns that collectively approximate the target transformation using deep learning-based training, we generated a complex-valued input-output image dataset for each target $\boldsymbol{A}$. To cover a wide range of spatial patterns, each input image in the dataset has a different sparsity ratio with randomly chosen pixel values. We also included rotated versions of each training image. We can summarize our input image dataset as

$$(4P + 8P + 16P + 32P + 48P + 64P) \times 4R \times S \tag{5.15}$$

where S refers to the number of images for each sub-category of the training image set defined by kP for $k \in \{4,8,16,32,48,64\}$, which indicates a training image where k pixels out of N_i pixels are chosen to be nonzero (with all the rest of the pixels being zero). Hence, k indicates the fill factor for a given image. We choose $S = 15{,}000$ for the training and $S = 7{,}500$ for the test image sets. Also, $4R$ in Equation 5.15 indicates the four different image rotations of a given training image, where the rotation angles are determined as $0°, 90°, 180°$ and $270°$. For example, $16P$ in Equation 5.15 indicates that randomly chosen 16 pixels out of 64 pixels of an image are nonzero and the remaining 48 pixels are zero. Following this formalism, we generated a total of $360K$ images for the training dataset and $180K$ for the test image dataset. Moreover, if a pixel in an image was chosen as nonzero, it took an independent random value from the set $\left\{\frac{32}{255}, \frac{33}{255}, \cdots, \frac{254}{255}, \frac{255}{255}\right\}$. Here the lower bound was chosen so that the "on" pixels can be well-separated from the zero-valued "off" pixels.

In this paper, we used the same input ($\boldsymbol{i}$) image dataset for all the transformation matrices ($\boldsymbol{A}$) that we utilized as our testbed. However, since the chosen linear transforms are different, the ground truth output fields are also different in each case, and were calculated based on $\boldsymbol{o} = \boldsymbol{A}\boldsymbol{i}$.

As discussed in Section 2.3, our forward model implements Equation 5.2 and the DFT operations are performed using the fast Fourier transform algorithm[179]. In our deep learning models, we chose the loss function as shown in Equation 5.10. All the networks were trained using Python (v3.6.5) and TensorFlow (v1.15.0, Google Inc.), where the Adam optimizer was

selected during the training. The learning rate, batch size and the number of training epochs were set to be 0.01, 8 and 50, respectively.

Computation of all-optical transformation performance metrics

As our all-optical transformation performance metrics, we used (i) the transformation error, (ii) the cosine similarity between the ground truth and the estimate transformation matrices, (iii) normalized output MSE and (iv) the mean output diffraction efficiency.

The first metric, the transformation error, is defined in Equation 5.5, which was used for both pseudoinverse-based diffractive designs and deep learning-based designs. The second chosen metric is the cosine similarity between two complex-valued matrices, which is defined as

$$U(\boldsymbol{a}, \widehat{\boldsymbol{a}}') = \frac{|\langle \boldsymbol{a}, \widehat{\boldsymbol{a}}' \rangle|}{\sqrt{\sum_{l=1}^{N_i N_o} |\boldsymbol{a}[l]|^2} \sqrt{\sum_{l=1}^{N_i N_o} |\widehat{\boldsymbol{a}}'[l]|^2}} = \frac{|\boldsymbol{a}^H \widehat{\boldsymbol{a}}'|}{\sqrt{\|\boldsymbol{a}\|^2} \sqrt{\|\widehat{\boldsymbol{a}}'\|^2}} \tag{5.16}$$

We use the notation $U(\boldsymbol{A}, \widehat{\boldsymbol{A}}')$ interchangeably with $U(\boldsymbol{a}, \widehat{\boldsymbol{a}}')$, both referring to Equation 5.16. Note that, even though the transformation error and cosine similarity metrics that are given by Equations 5.5 and 5.16, respectively, are related to each other, they end up with different quantities. The relationship between these two metrics can be revealed by rewriting Equation 5.5 as

$$\|\boldsymbol{a} - \widehat{\boldsymbol{a}}'\|^2 = (\boldsymbol{a} - \widehat{\boldsymbol{a}}')^H (\boldsymbol{a} - \widehat{\boldsymbol{a}}') = \|\boldsymbol{a}\|^2 + \|\widehat{\boldsymbol{a}}'\|^2 - 2Re\{\boldsymbol{a}^H \widehat{\boldsymbol{a}}'\} \tag{5.17}$$

where $Re\{\cdot\}$ operator extracts the real part of its input. As a result, apart from the vector normalization constants, $\|\boldsymbol{a}\|^2$ and $\|\widehat{\boldsymbol{a}}'\|^2$, Equations 5.16 and 5.17 deal with the magnitude and real part of the inner product ($\boldsymbol{a}^H\widehat{\boldsymbol{a}}'$), respectively.

For the third metric, the normalized MSE calculated at the output FOV, we used the following equation:

$$E[\|\widetilde{\boldsymbol{o}} - \widetilde{\boldsymbol{o}}'\|^2] = \frac{1}{S_T N_o}\sum_{s=1}^{S_T}\sum_{l=1}^{N_o}|\widetilde{\boldsymbol{o}}_s[l] - \widetilde{\boldsymbol{o}}'_s[l]|^2 = \frac{1}{S_T N_o}\sum_{s=1}^{S_T}\sum_{l=1}^{N_o}|\sigma_s \boldsymbol{o}_s[l] - \sigma'_s \boldsymbol{o}'_s[l]|^2 \tag{5.18}$$

where $E[\cdot]$ is the expectation operator and S_T is the total number of the image samples in the test dataset. The vectors $\boldsymbol{o}_s = \boldsymbol{A}\boldsymbol{i}_s$ and $\boldsymbol{o}'_s = \boldsymbol{A}'\boldsymbol{i}_s$ represent the ground truth and the estimated output fields (at the output FOV), respectively, for the s^{th} input image sample in the dataset, $\boldsymbol{i}_s$. The normalization constant, σ_s is given by Equation 5.12, and σ'_s, can be computed from Equation 5.13, by replacing σ'_s and $\boldsymbol{o}'_s$ by $\sigma'_{s,c}$ and $\boldsymbol{o}'_{s,c}$, respectively.

Finally, we chose the mean diffraction efficiency of the diffractive system as our last performance metric, which is computed as

$$E\left[\frac{\|\boldsymbol{o}'\|^2}{\|\boldsymbol{i}\|^2}\right] = \frac{1}{S_T}\sum_{s=1}^{S_T}\frac{\sum_{l=1}^{N_o}|\boldsymbol{o}'_s[l]|^2}{\sum_{l=1}^{N_i}|\boldsymbol{i}_s[l]|^2} \tag{5.19}$$

Random generation of ground truth matrices

To create the unitary transformations, as presented in Fig. 5.1.b, we first generated a complex-valued Givens rotation matrix, which is defined for a predetermined $i, j \in \{1,2, \cdots, N_i\}$ and $i \neq j$ pair as

$$\boldsymbol{R}_{ij}[n,m] = \begin{cases} 1, & \text{if } n = m, n \neq i \text{ and } n \neq j \\ e^{j\theta_1}\cos\theta_3, & \text{if } n = m = i \\ e^{-j\theta_1}\cos\theta_3, & \text{if } n = m = j \\ e^{j\theta_2}\sin\theta_3, & \text{if } n = i \text{ and } m = j \\ -e^{-j\theta_2}\sin\theta_3, & \text{if } n = j \text{ and } m = i \\ 0, & \text{otherwise} \end{cases} \tag{5.20}$$

where $\theta_1, \theta_2, \theta_3 \in [0,2\pi)$ are *randomly* generated phase values. Then a unitary matrix was computed as

$$\boldsymbol{R} = \prod_{t=1}^{T} \boldsymbol{R}_{i_t j_t} \tag{5.21}$$

where (i_t, j_t) pair is randomly chosen for each *t*. We used $T = 10^5$ in our simulations. As a result, for each *t* in Equation 21, (i_t, j_t) and $(\theta_1, \theta_2, \theta_3)$ were chosen randomly. It is straightforward to show that the resulting $\boldsymbol{R}$ matrix in Equation 21 is a unitary matrix.

To compute the nonunitary but invertible matrices, we first generated two unitary matrices $\boldsymbol{R}_U$ and $\boldsymbol{R}_V$, as described by Equations 5.20 and 5.21, and then a diagonal matrix $\boldsymbol{X}$. The diagonal elements of $\boldsymbol{X}$ takes uniformly, independently and identically generated random real values in the range $[0.3,1]$, where the lower limit is determined to be large enough to prevent

numerical instabilities and the upper limit is determined to prevent amplification of the orthonormal components of $\boldsymbol{R}_U$ and $\boldsymbol{R}_V$. Then, the nonunitary but invertible matrix is generated as $\boldsymbol{R}_U \boldsymbol{X} \boldsymbol{R}_V^H$, which is in the form of the SVD of the resulting matrix. It is straightforward to show that the resulting matrix is invertible. However, to make sure that it is nonunitary, we numerically compute its Hermitian and its inverse separately, and confirm that they are not equal. Similarly, to compute the noninvertible transformation matrix, as shown in e.g., Fig. 5.4.b, we equated the randomly chosen half of the diagonal elements of $\boldsymbol{X}$ to zero and randomly chose the remaining half to be in the interval [0.3,1]. Following this, we computed the noninvertible matrix as $\boldsymbol{R}_U \boldsymbol{X} \boldsymbol{R}_V^H$, by re-computing new unitary matrices $\boldsymbol{R}_U$ and $\boldsymbol{R}_V$, which end up to be completely different from the $\boldsymbol{R}_U$ and $\boldsymbol{R}_V$ matrices that were computed for the nonunitary and invertible transform.

The 2D DFT operation for the square input aperture located at the center of the input plane was defined by

$$\boldsymbol{o_{2D}}[p,q] = \frac{1}{\sqrt{N_i}} \sum_{n=-\frac{\sqrt{N_i}}{2}}^{\frac{\sqrt{N_i}}{2}-1} \sum_{m=-\frac{\sqrt{N_i}}{2}}^{\frac{\sqrt{N_i}}{2}-1} \boldsymbol{i_{2D}}[n,m] e^{-j\frac{2\pi}{\sqrt{N_i}}(pn+qm)} \tag{5.22}$$

where $\boldsymbol{i_{2D}}$ and $\boldsymbol{o_{2D}}$ represent the 2D fields on the input and output FOVs, respectively, and $n,m,p,q \in \left\{-\frac{\sqrt{N_i}}{2}, -\frac{\sqrt{N_i}}{2}+1, \cdots, \frac{\sqrt{N_i}}{2}-1\right\}$. Here we assume that the square-shaped input and output FOVs have the same area and number of pixels, i.e., $N_i = N_o$. Moreover, since we assume that the input and output FOVs are located at the center of their associated planes, the space and

frequency indices start from $-\sqrt{N_i}/2$. Therefore, the $\boldsymbol{A}$ matrix associated with the 2D centered DFT, which is shown in Fig. 5.7.b, performs the transform given in Equation 5.22.

The permutation ($\boldsymbol{P}$) operation performs a one-to-one mapping of the complex-value of each pixel on the input FOV onto a different location on the output FOV. Hence the randomly selected transformation matrix ($\boldsymbol{A} = \boldsymbol{P}$) associated with the permutation operation has only one nonzero element along each row, whose value equals to 1, as shown in Fig. 5.10b.

Finally, the transformation matrix corresponding to the high-pass filtering operation, as shown in Fig. 5.13b, is generated from the Laplacian high-pass filter whose 2D convolution kernel is

$$\begin{bmatrix} 1 & 4 & 1 \\ 4 & -20 & 4 \\ 1 & 4 & 1 \end{bmatrix} \tag{5.23}$$

After generating the 2D matrix by applying the appropriate vectorization operation, we also normalize the resulting matrix with its largest singular value, to prevent the amplification of the orthonormal components.

Penalty term for improved diffraction efficiency

To increase the diffraction efficiency at the output FOV of a diffractive network design, we used the following modified loss function:

$$L = c_M L_M + c_D L_D \tag{5.24}$$

where L_M is the MSE loss term which is given in Equation 5.10 and L_D is the additional loss term that penalizes poor diffraction efficiency:

$$L_D = e^{-\alpha X} \tag{5.25}$$

where X is the diffraction efficiency term which is given by Equation 5.19. In Equations 5.24 and 5.25, c_M, c_D and α are the user-defined weights. In earlier designs where the diffraction efficiency has not been penalized or taken into account during the training phase, c_M and c_D were taken as 1 and 0, respectively.

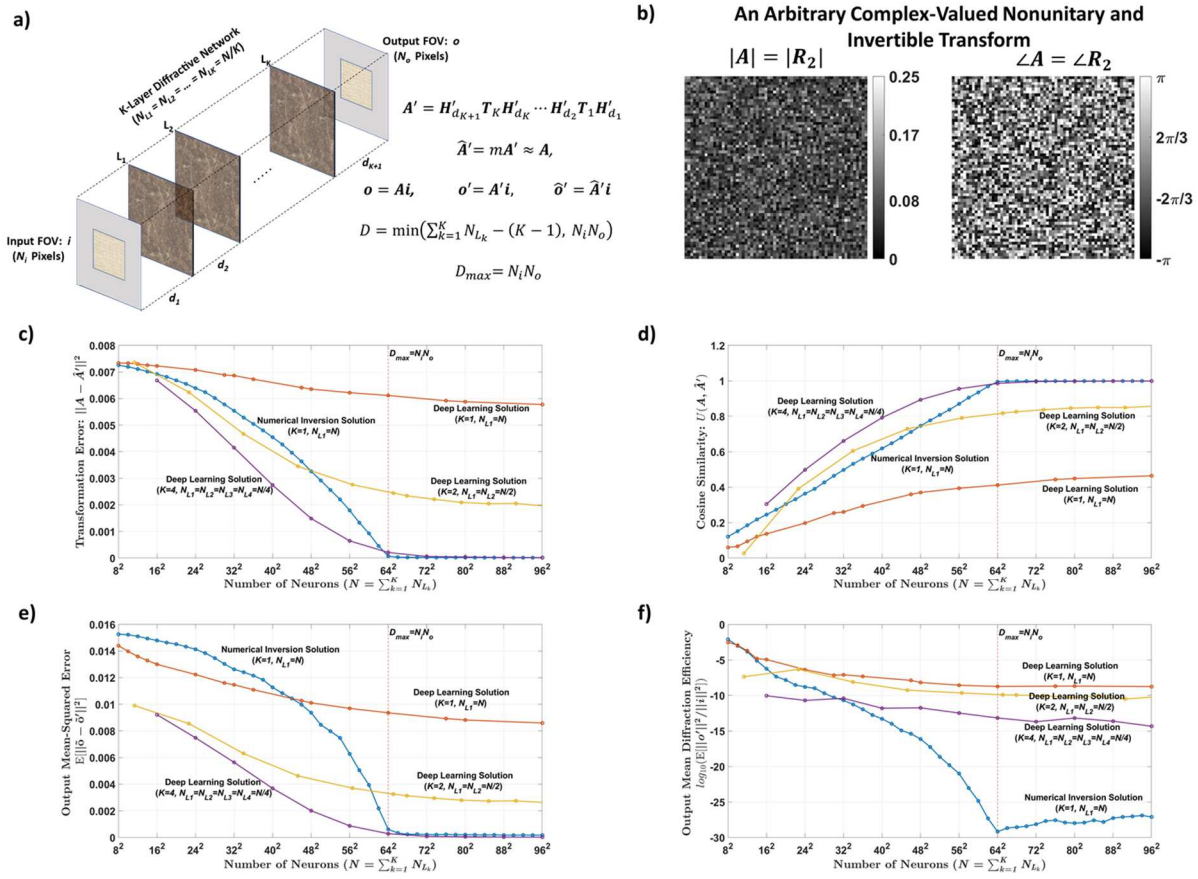

Fig. 5.4 Diffractive all-optical transformation results for an arbitrary complex-valued nonunitary and invertible transform. Follows the caption of Fig. 5.1.

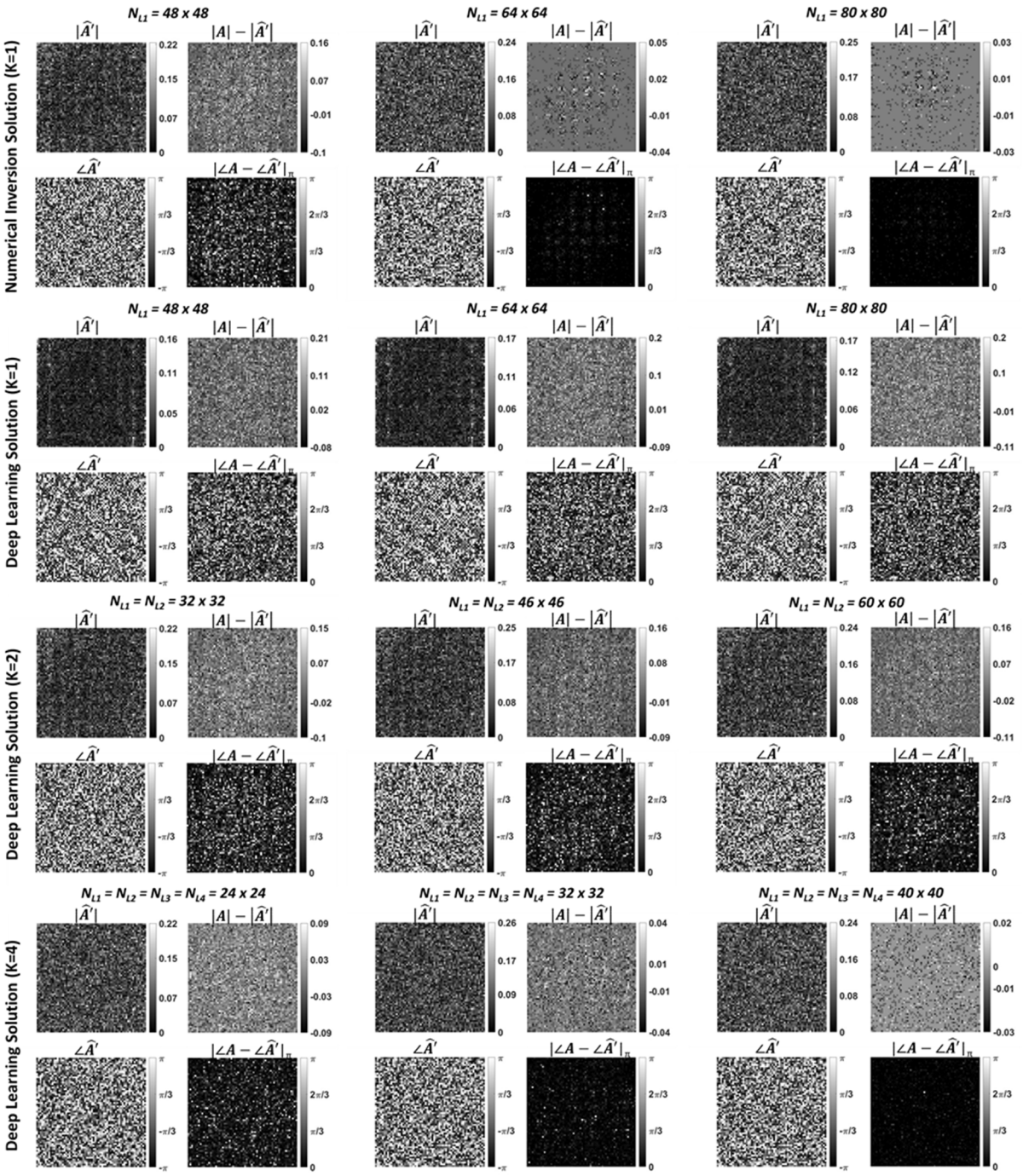

Fig. 5.5 Diffractive all-optical transformations and their differences from the ground truth, target transformation (A) where, $A = R_2$, is an arbitrary complex-valued nonunitary and invertible transform. Follows the caption of Fig. 5.2.

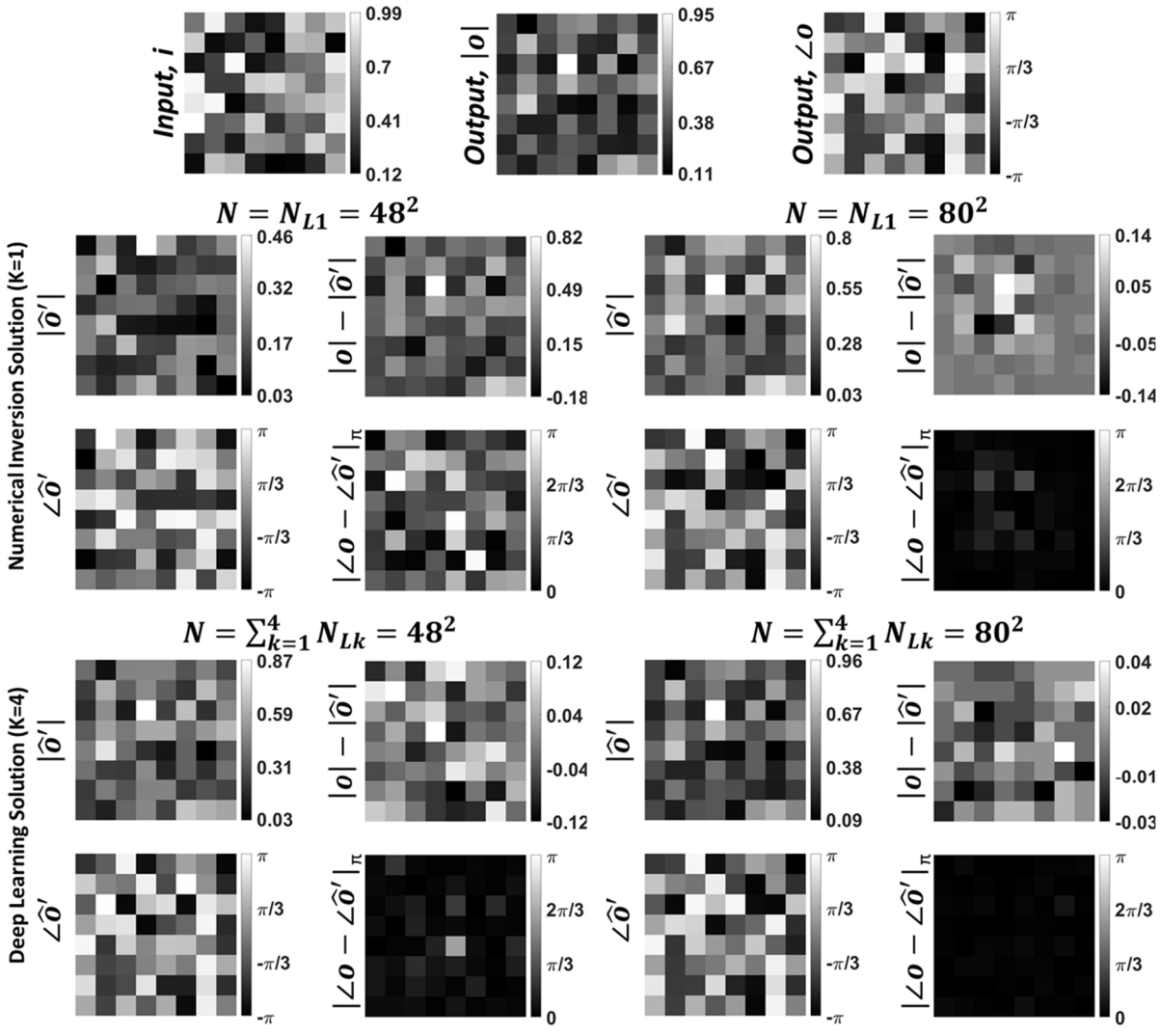

Fig. 5.6 Sample input-output images for the ground truth transformation presented in Fig. 5.4b and the optical outputs by the diffractive designs for two different choices of N ($N = 48^2$ and $N = 80^2$). Follows the caption of Fig. 5.3.

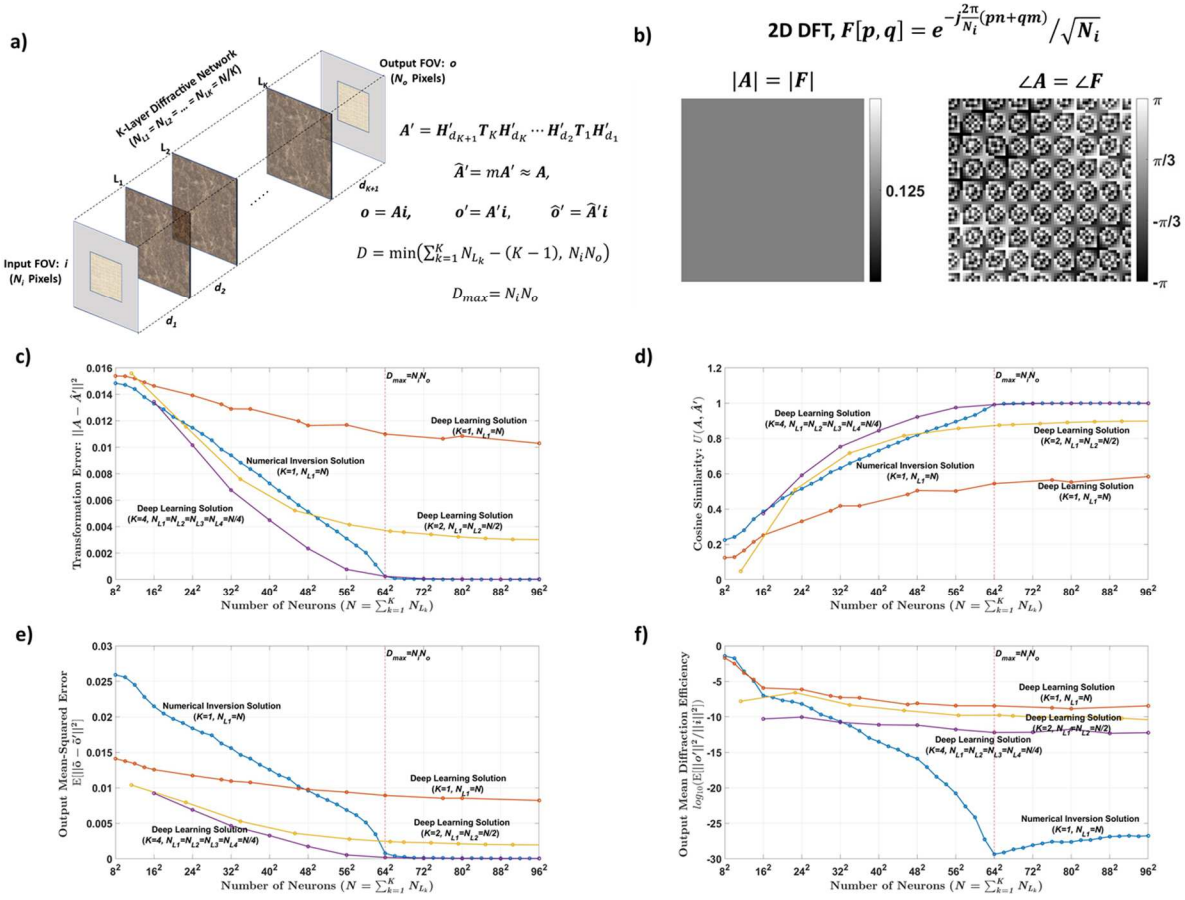

Fig. 5.7 Diffractive all-optical transformation results for 2D discrete Fourier transform. Follows the caption of Fig. 5.1.

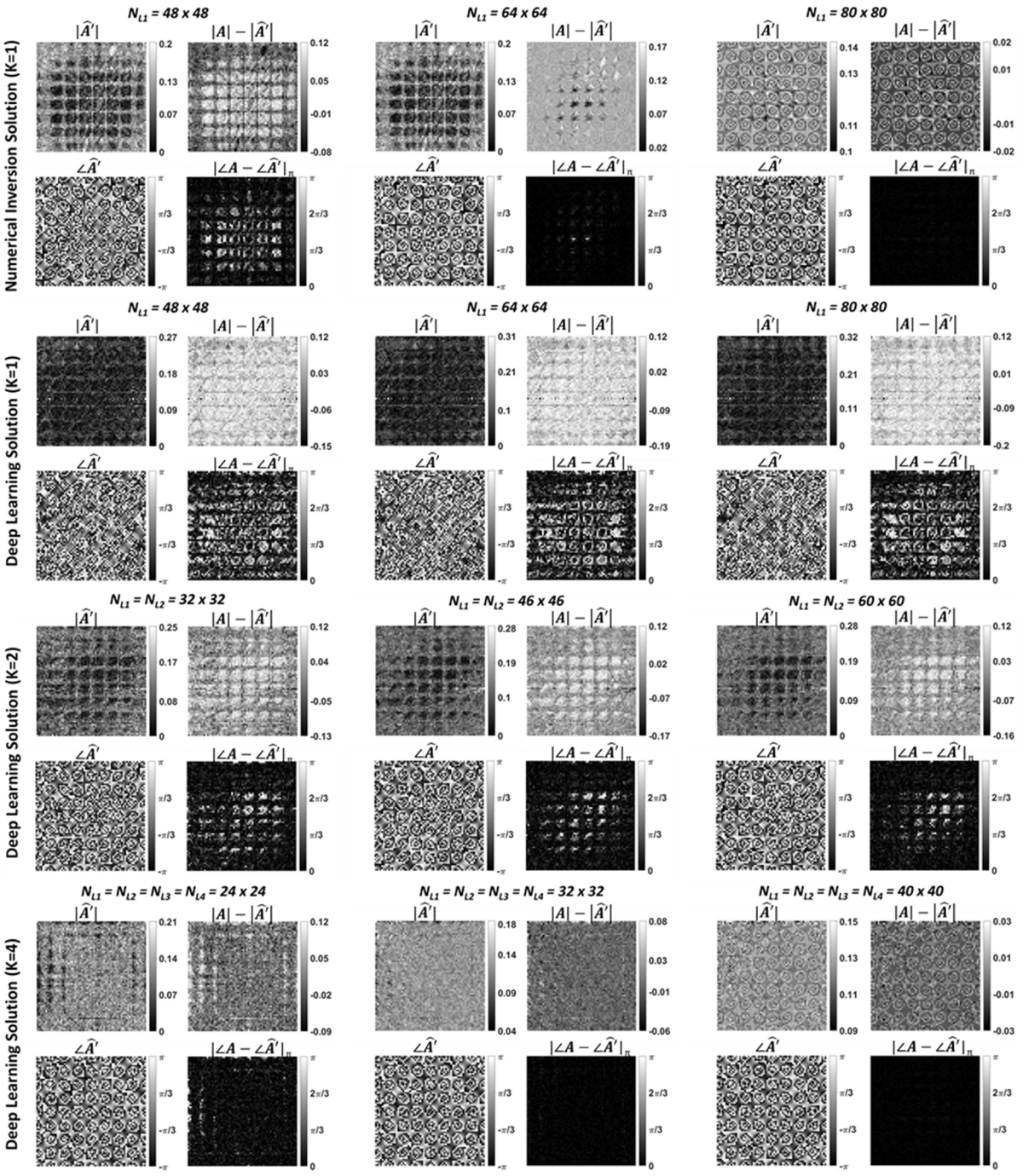

Fig. 5.8 Diffractive all-optical transformations and their differences from the ground truth, target transformation (A) where, $A = F$, represents 2D discrete Fourier transform. Follows the caption of Fig. 5.2.

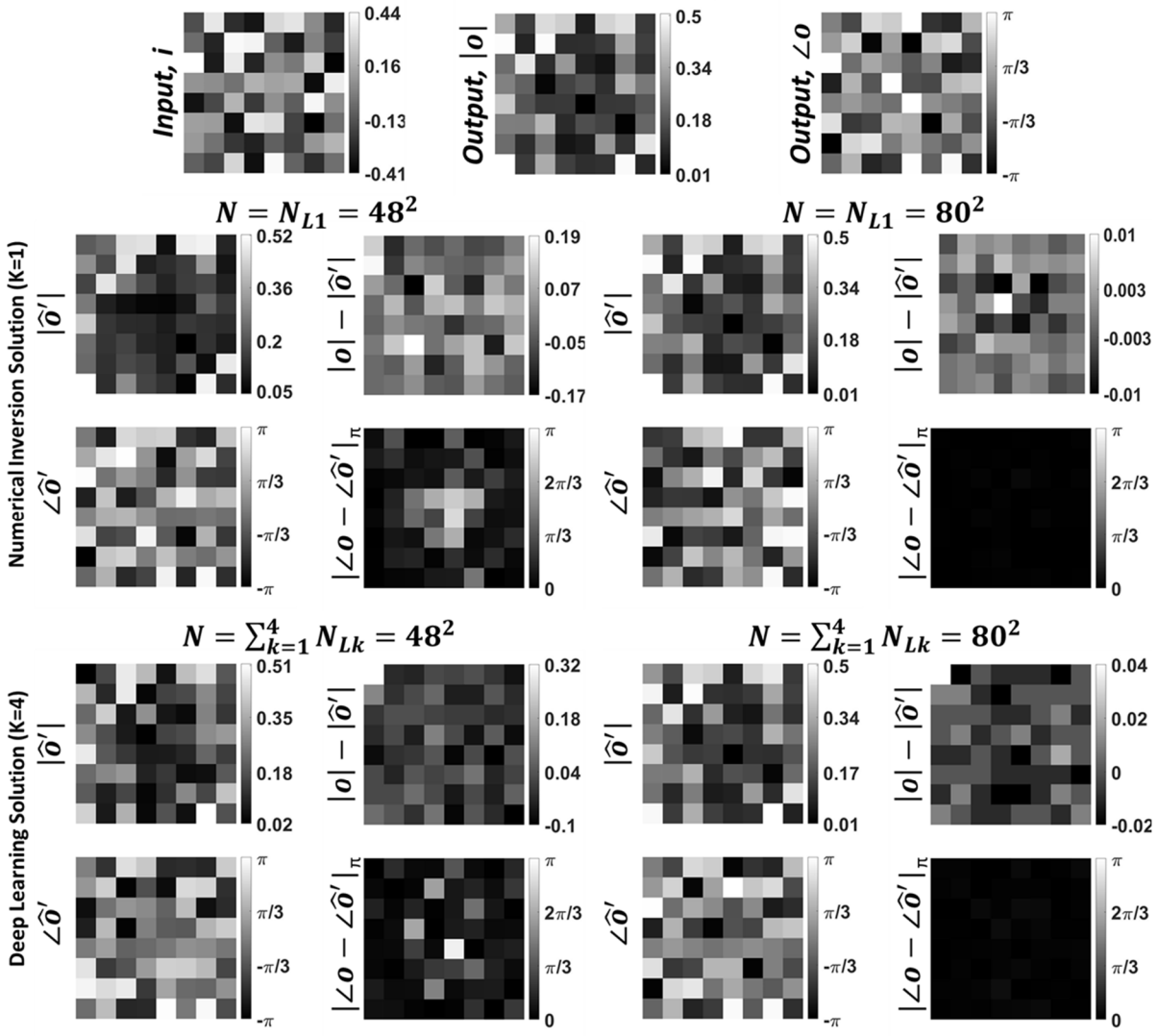

Fig. 5.9 Sample input-output images for the ground truth transformation presented in Fig. 5.7b and the optical outputs by the diffractive designs for two different choices of N ($N = 48^2$ and $N = 80^2$). Follows the caption of Fig. 5.3.

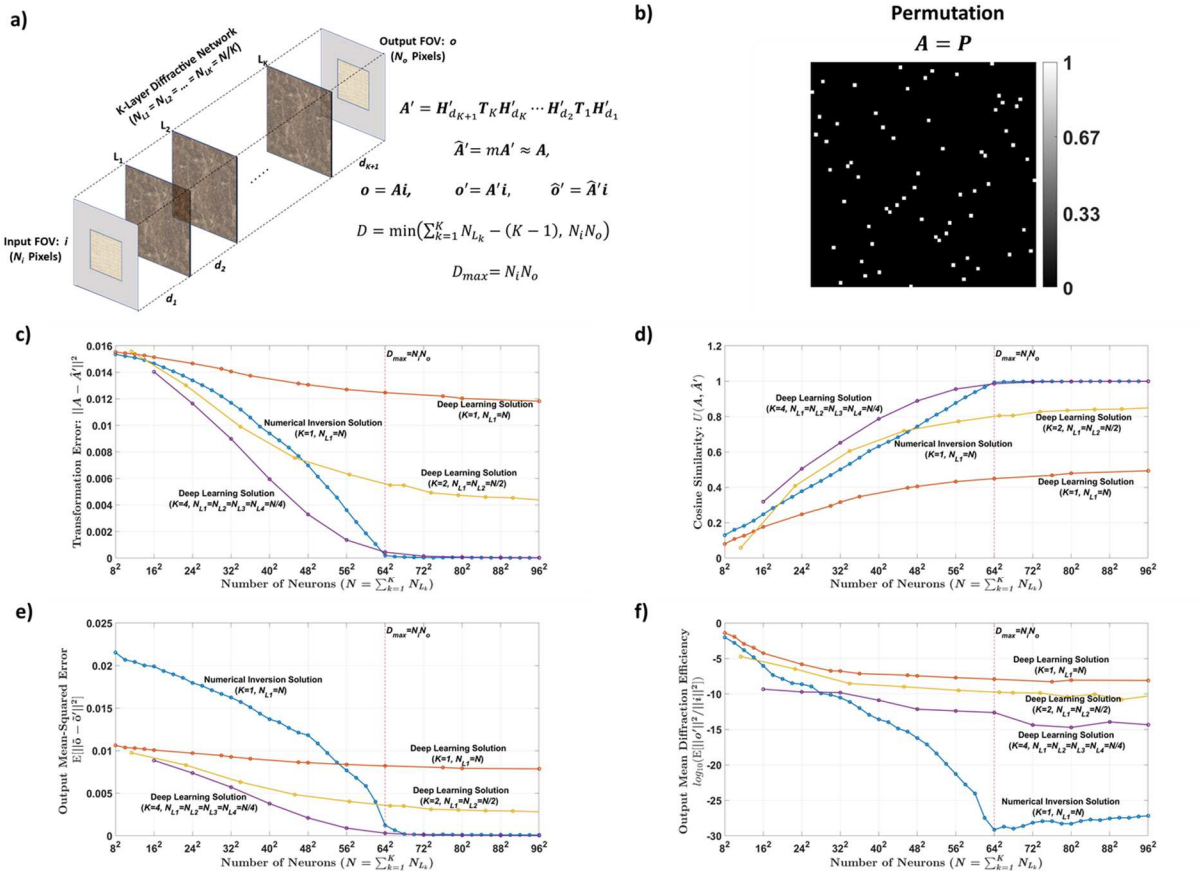

Fig. 5.10 Diffractive all-optical transformation results for an arbitrary permutation matrix, $A = P$. Follows the caption of Fig. 5.1.

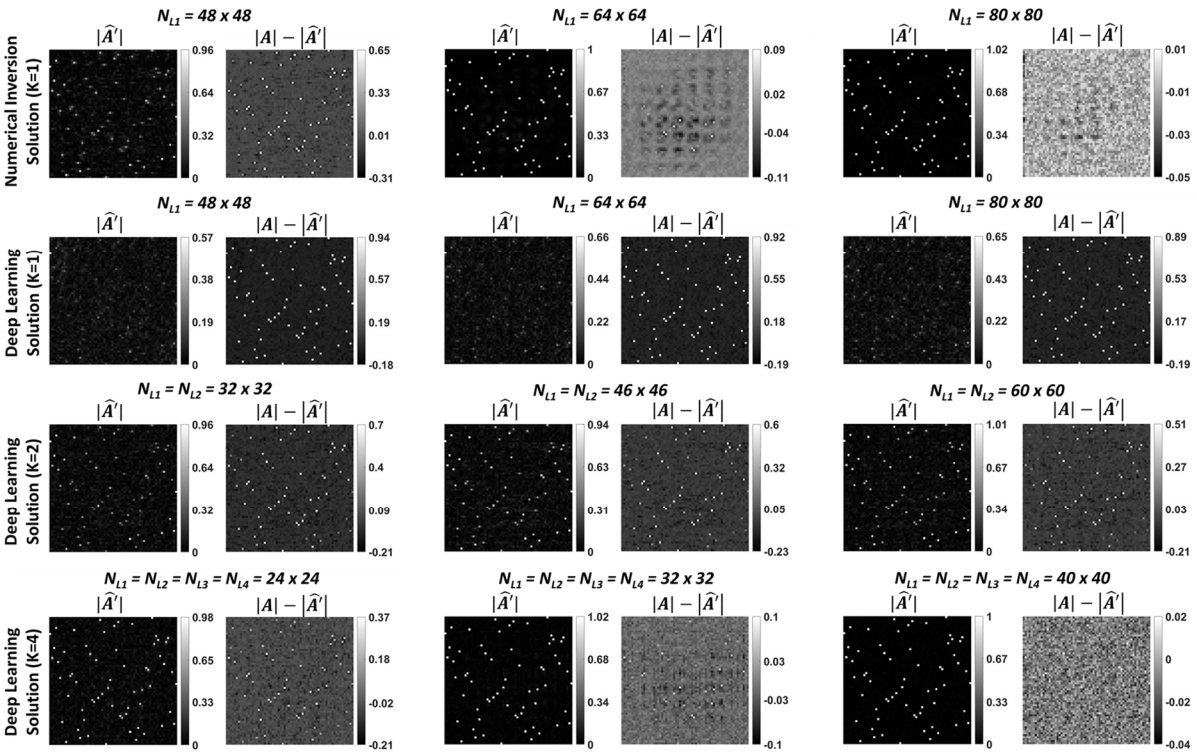

Fig. 5.11 Diffractive all-optical transformations and their differences from the ground truth, target transformation (A) where, $A = P$, represents an arbitrary permutation matrix. Follows the caption of Fig. 5.2.

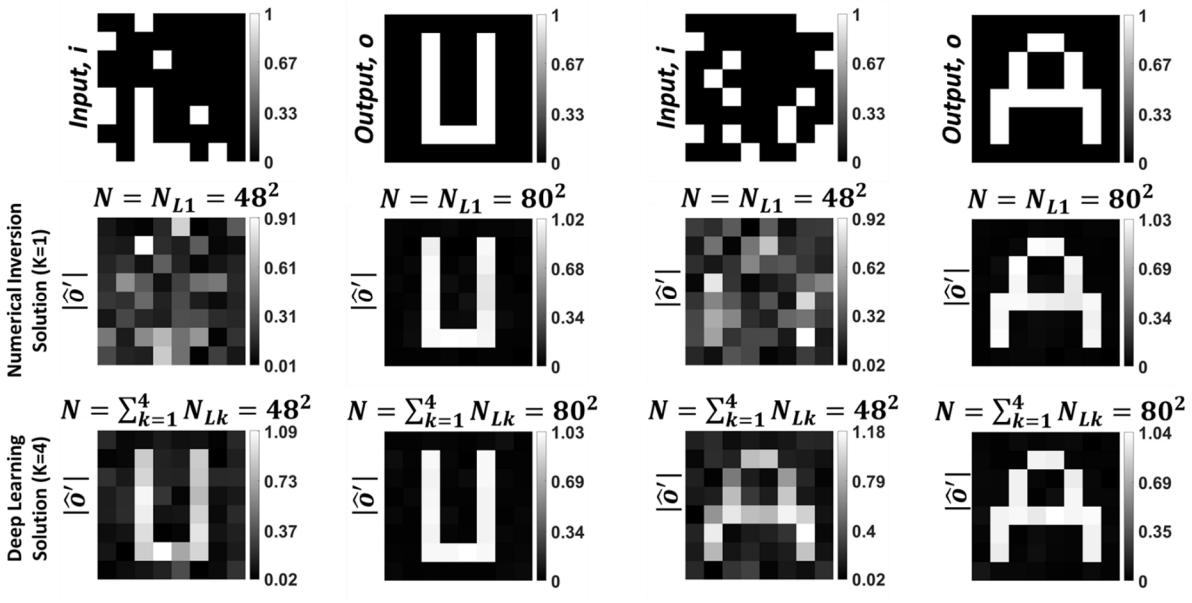

Fig. 5.12 Sample input-output images for the ground truth transformation presented in Fig. 5.10b and the optical outputs by the diffractive designs for two different choices of N ($N = 48^2$ and $N = 80^2$). Follows the caption of Fig. 5.3.

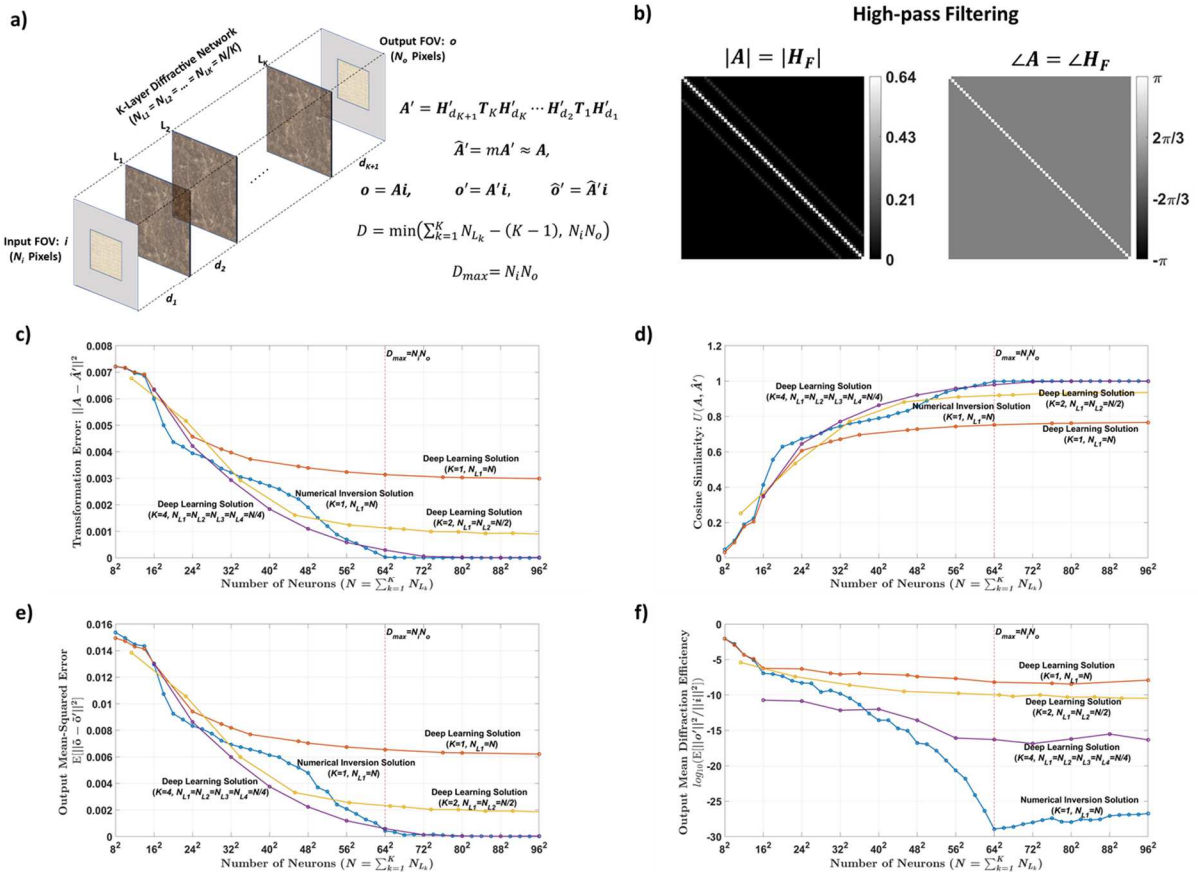

Fig. 5.13 Diffractive all-optical transformation results for a high-pass filtered imaging operator, $A = H_F$.

Follows the caption of Fig. 5.1.

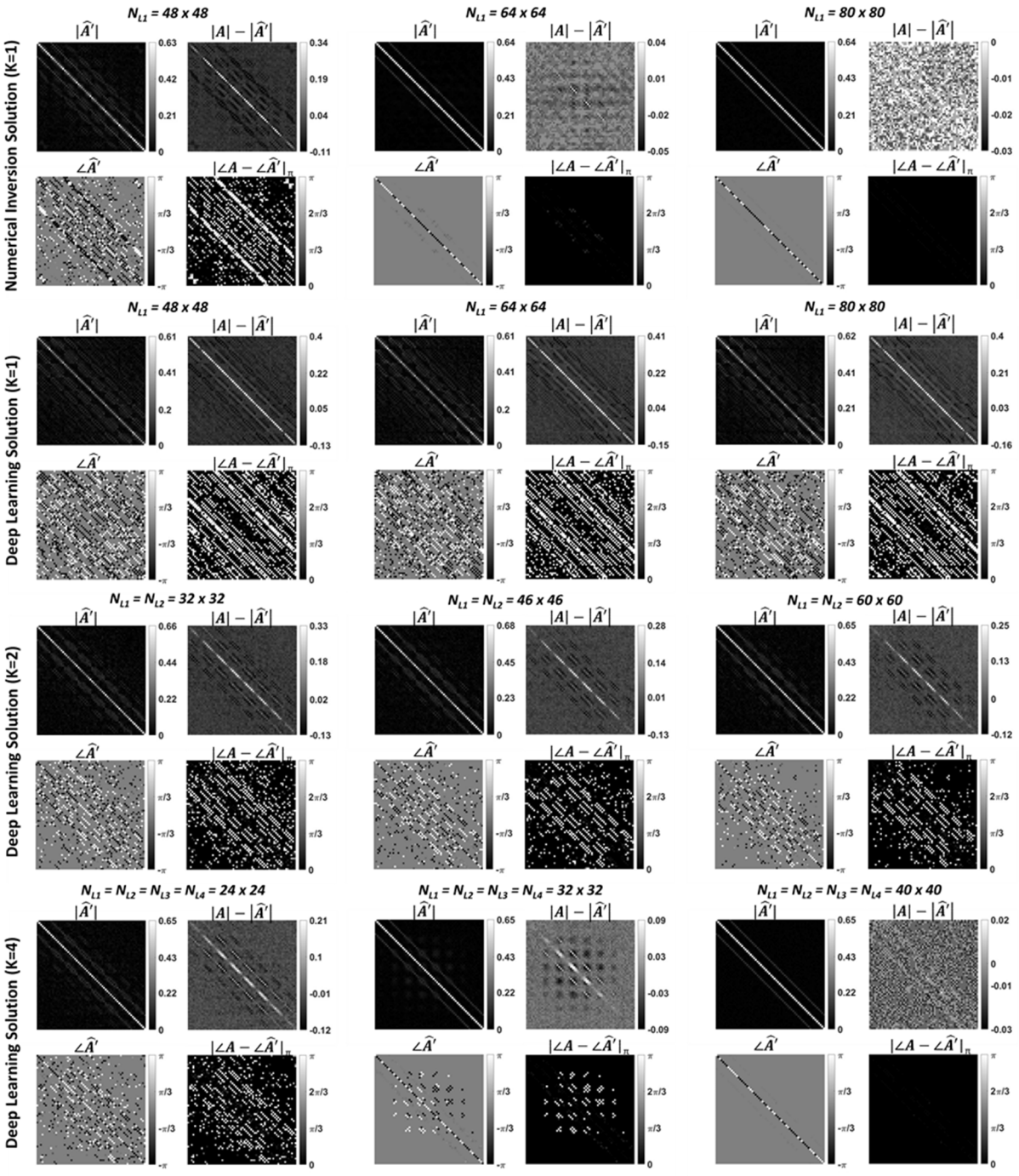

Fig. 5.14 Diffractive all-optical transformations and their differences from the ground truth, target transformation (A) where, $A = H_F$, represents a high-pass filtered imaging operator. Follows the caption of Fig. 5.2.

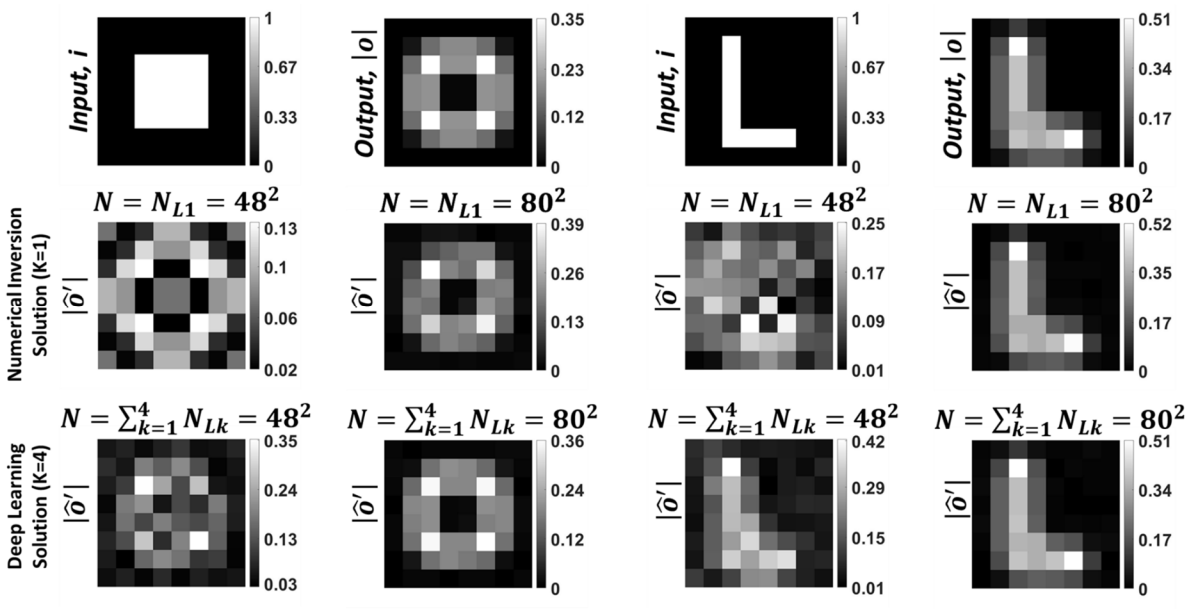

Fig. 5.15 Sample input-output images for the ground truth transformation presented in Fig. 5.13b and the optical outputs by the diffractive designs for two different choices of N ($N = 48^2$ and $N = 80^2$). Follows the caption of Fig. 5.3.

Chapter 6 Diffractive Interconnects: All-Optical Permutation Operation Using Diffractive Networks

Parts of this chapter have previously been published in D. Mengu et al. “Diffractive interconnects: all-optical permutation operation using diffractive networks”, Nanophotonics. DOI: 10.1515/nanoph-2022-0358. This chapter presents the experimental demonstration of diffractive permutation networks by extending the findings of the previous study to intensity-to-intensity transformations.

Permutation matrices form an important computational building block frequently used in various fields including e.g., communications, information security and data processing. Optical implementation of permutation operators with relatively large number of input-output interconnections based on power-efficient, fast, and compact platforms is highly desirable. Here, we present diffractive optical networks engineered through deep learning to all-optically perform permutation operations that can scale to hundreds of thousands of interconnections between an input and an output field-of-view using passive transmissive layers that are individually structured at the wavelength scale. Our findings indicate that the capacity of the diffractive optical network in approximating a given permutation operation increases proportional to the number of diffractive layers and trainable transmission elements in the system. Such deeper diffractive network designs can pose practical challenges in terms of physical alignment and output diffraction efficiency of the system. We addressed these challenges by designing misalignment tolerant diffractive designs that can all-optically perform arbitrarily-selected permutation operations, and experimentally demonstrated, for the first time, a diffractive permutation network that operates at THz part of the spectrum. Diffractive permutation networks

might find various applications in e.g., security, image encryption and data processing, along with telecommunications; especially with the carrier frequencies in wireless communications approaching THz-bands, the presented diffractive permutation networks can potentially serve as channel routing and interconnection panels in wireless networks.

6.1 Introduction

Permutation is one of the basic computational operations that has played a key role in numerous areas of engineering e.g., computing[181], communications[182], encryption[183], data storage[184], remote sensing[185] and data processing[136]. Historically, electronic integrated circuits have been the established implementation medium for the permutation operation and other space-variant linear transformations, while the research on optical computing has been mainly focused on using the Fourier transform approximation of thin lenses covering various applications in space-invariant transformations e.g., convolution/correlation. On the other hand, as photonic switching devices and optical waveguide technology have become the mainstream communication tools on high-end applications e.g., fiber optic networks, supercomputers and data centers, various approaches have been developed towards all-optical implementation of permutation operation and other space-variant transformations based on e.g., Mach-Zehnder interferometers[186], optical switches[187], photonic crystals[188], holographically recorded optical elements[189–191], off-axis lenslet arrays[192,193] and arrays of periodic grating-microlens doublets[194]. The development of compact, low-power optical permutation and interconnection devices can have significant impact on next-generation communication systems e.g., 6G networks[195,196], as well as other applications such as optical data storage[197] and image encrypting cameras[198–200].

With the widespread availability of high-end graphics processing units (GPU) and the massively growing amounts of data, the past decade has witnessed major advances in deep learning, dominating the field of digital information processing for various engineering applications including e.g., image segmentation and classification[46,201–203], natural language processing[204,205], among others[206]. The statistical inference and function approximation capabilities of deep neural networks have also been exploited to produce state-of-the-art

performance for computational inverse problems in many imaging and sensing applications including e.g., microscopy[62,64,93,207–210], quantitative phase imaging[50–52,211–213] and others[44,47,58,63,214–219]. Beyond these data processing tasks, deep learning can also provide task-specific solutions to challenging inverse optical design problems for numerous applications including nanophotonics[67,68], metamaterials[220], imaging and sensing[221–226]. However, as the success and the applications of deep learning grow further, the electronic parallel computing platforms e.g., GPUs, hosting deep neural networks and other machine learning algorithms have started to bring some limitations due to their power- and bandwidth-hungry operation. Moreover, the pace of the advances in computational capacity of the integrated circuits has fallen behind the exponential increase predicted by the Moore's law[227]. These factors have fueled a tremendous amount of effort towards the development of optical machine learning schemes and other photonic computing devices that can partially reduce the computational burden on the electronics leading to power-efficient, massively parallel, high-speed machine learning systems. While most of the arising optical computing techniques rely on integrated photonic devices and systems compatible with the integrated waveguide technology[20,69,70,74,228–230], an alternative option towards exploiting photons for machine learning and the related computing tasks is to use complex modulation media and free-space light propagation and diffraction, which is particularly suitable for visual computing applications where the information is already carried by the optical waves (e.g., of a scene or target object) in free-space[138].

Motivated by these pressing needs, Diffractive Deep Neural Networks (D^2NN)[77,172] have emerged as an optical machine learning framework that utilizes deep learning to engineer light-matter interactions over a series of diffractive surfaces so that a desired statistical inference or deterministic computing task is realized all-optically as the light waves propagate through

structured surfaces. According to this framework, the physical parameters determining the phase and/or amplitude of light over each independently controllable unit, i.e., the 'diffractive neuron', are updated through the stochastic gradient descent and error-backpropagation algorithms based on a loss function tailored specifically for a given task. The weights of the connections between the diffractive neurons/features on successive layers, on the other hand, are dictated by the light diffraction in free-space. Once the deep learning-based training, which is a one-time effort, is completed using a computer, the resulting transmissive/reflective diffractive layers are fabricated using e.g., lithography or 3D printing, to physically form the diffractive network that *completes* a given inference or computational task at the speed of light using entirely passive modulation surfaces, offering a task-specific, power-efficient and fast optical machine learning platform.

Based on the D^2NN framework, here we demonstrate diffractive optical network designs that were trained to all-optically perform a given permutation operation between the optical intensities at the input and output fields-of-view, capable of handling hundreds of thousands of interconnections with diffraction limited resolution. We quantified the success of the presented diffractive optical networks in approximating a given, randomly-selected permutation operation as a function of the number of diffractive neurons and transmissive layers used in the diffractive network design. We also laid the foundations toward practical implementations of diffractive permutation networks by investigating the impact of various physical error sources, e.g., lateral and axial misalignments and unwanted in-plane layer rotations, on the quality/accuracy of the optically-realized interconnection weights and the permutation operation. Moreover, we showed that the diffractive optical permutation networks can be trained to be resilient against possible misalignments as well as imperfections in the diffractive layer fabrication and assembly. Finally, we present the first proof-of-concept experimental demonstration of diffractive permutation

networks by all-optically achieving a permutation matrix of size 25×25, effectively realizing 625 interconnections based on 3D-printed diffractive layers operating at the THz part of the spectrum.

The presented diffractive optical permutation networks can readily find applications in THz-band communication systems serving as communication channel patch panels; furthermore, the underlying methods and design principles can be broadly extended to operate at other parts of the electromagnetic spectrum, including the visible and IR wavelengths, by scaling each diffractive feature size proportional to the wavelength of light[231] and can be used for image encryption in security cameras[232] and optical data storage systems, among other applications[233–236].

6.1 Results

Figure 6.1 illustrates the presented free-space permutation interconnect concept designed around diffractive optical networks using the D^2NN framework. As shown in Fig. 6.1, the presented permutation interconnect scheme does not use any standard optical components such as lenses, and instead relies on a series of passive, phase-only diffractive surfaces. Due to the passive nature of these layers, the diffractive optical network shown in Fig. 6.1 does not consume any power except for the illumination light, providing a power-efficient permutation operation in a compact footprint of ~600λ along the longitudinal axis), which could be further squeezed as needed. The 5-layer diffractive optical permutation network design shown in Fig. 6.1 was trained through supervised learning to all-optically realize a desired permutation operation, $\boldsymbol{P}$, between the light intensity signals at the input and output FOVs, each with $N_i = N_o = 400$ (20×20) individual pixels of size 2λ×2λ. Stated differently, this permutation operation controls in total of $N_i N_o = 0.16$ million optical intensity connections.

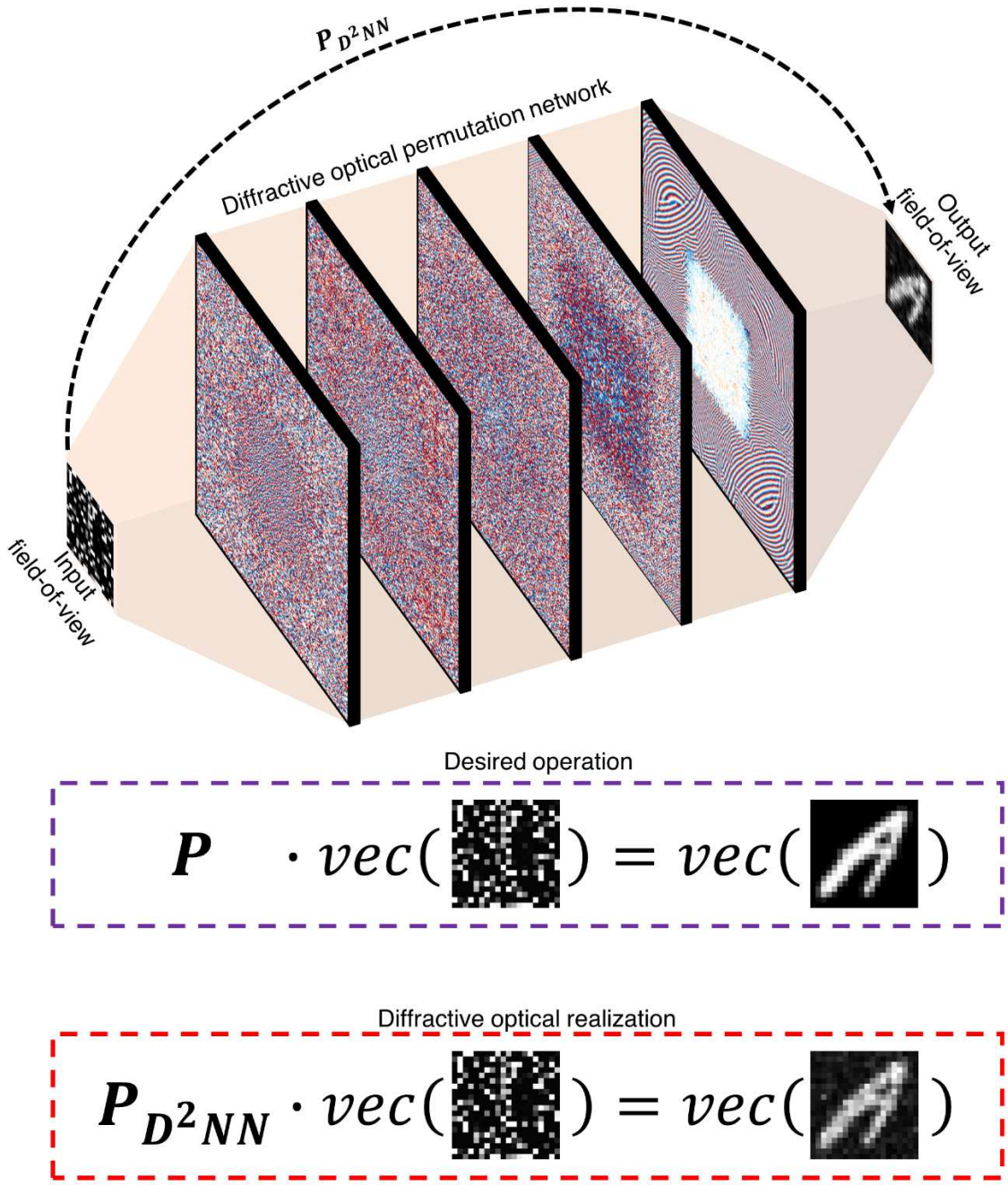

Fig. 6.1 The schematic of a 5-layer diffractive permutation network, all-optically realizing 0.16 million interconnects between an input and output field-of-view. The presented diffractive permutation network was trained to optically realize an arbitrarily-selected permutation operation between the light intensities over N_i=400=20×20 input and N_o=400=20×20 output pixels, establishing total 0.16 million desired interconnections based on 5 phase-only diffractive layers, each containing 40K (200×200) diffractive neurons/features.

The supervised nature of the training process of the diffractive permutation network necessitates the use of a set of input-output signal pairs (examples that satisfy $\boldsymbol{P}$) to compute a penalty term and the associated gradient-based updates with respect to the physical parameters of each diffractive neuron at every iteration. We set the optical signal of interest at the input and the output of the diffractive permutation scheme to be the light intensity, and as a result, the deep learning-based evolution of the presented diffractive permutation network shown in Fig. 6.1 was driven based on the mean-squared error (MSE) (see Methods section) between the ground-truth and the all-optically synthesized output intensity patterns at a given iteration. Since this loss function acts only on the light intensity, the diffractive optical network can enjoy an output phase-freedom in synthesizing the corresponding transformed optical intensity patterns within the output field-of-view. The light intensity, I, is related to the complex-valued field, U, through a nonlinear operation, $I = |U|^2$. If a pair of input-output complex-fields exists for a given diffractive network, i.e., $\{U_{in}, U_{out}\}$ and $\{U'_{in}, U'_{out}\}$, then the input field $U''_{in} = \alpha U_{in} + \beta U'_{in}$ will create the output field $U''_{out} = \alpha U_{out} + \beta U'_{out}$ at the output plane. In terms of the associated intensities, however, this direct linear extension does not hold since $|\alpha U_{out}|^2 + |\beta U'_{out}|^2 \neq |U''_{out}|^2$, making it challenging (in terms of the data generalization capability) to design diffractive optical networks for achieving a general purpose intensity-to-intensity transformation such as a permutation operation. To overcome this generalization challenge, we trained our diffractive permutation networks using ~4.7 million randomly generated input/output intensity patterns that satisfy the desired $\boldsymbol{P}$, instead of a standard benchmark image dataset (see the Methods).

After the training phase, we blindly tested each diffractive permutation network with test inputs that were never used during the training. Figure 6.2 illustrates 6 different randomly

generated blind testing inputs along with the corresponding all-optically permuted output light intensities. In the first two randomly generated input patterns shown in Fig. 6.2A, there is light coming out of all the input pixels/apertures at different levels of intensity. In the next two test input patterns shown in Fig. 6.2A, on the other hand, nearly half of the input apertures have nonzero light intensity and finally, the last two test inputs contain only 10 and 11 pixels/apertures with light propagating towards the 1st layer of the diffractive permutation network. When tested on 20K randomly generated blind testing input intensity patterns with different sparsity levels, the 5-layer diffractive permutation network shown in Fig. 6.1 achieves 18.61 dB peak-signal-to-noise ratio (PSNR), very well matching the ideal output response. For this randomly generated 20K testing data, Figure 6.2B also shows the distribution of PSNR as a function of the number of input pixels with nonzero light intensity, which reveals that the diffractive permutation network can permute relatively sparser inputs with a higher output image quality, achieving a PSNR of 25.82 dB.

In addition to randomly generated blind testing inputs, we further tested the diffractive permutation network shown in Fig. 6.1 on 18.75K EMNIST images; note that this diffractive network was trained only using randomly generated input/output intensity patterns that satisfy $\boldsymbol{P}$ and the EMNIST images constitute not only blind testing set but also a significant deviation from the statistical distribution of the training images. The input field-of-view contains the permuted EMNIST images ($\boldsymbol{P}^{-1}$) and the diffractive network inverts that permutation by all-optically performing $\boldsymbol{P}$ to recover the original images at the output plane (see Fig. 6.2). The performance of the diffractive permutation network was quantified based on both PSNR and Structural Similarity Index Measure (SSIM). With N_L=40K diffractive neurons on each layer, the 5-layer diffractive permutation network shown in Fig. 6.1 provides 19.18 dB and 0.85 for PSNR and

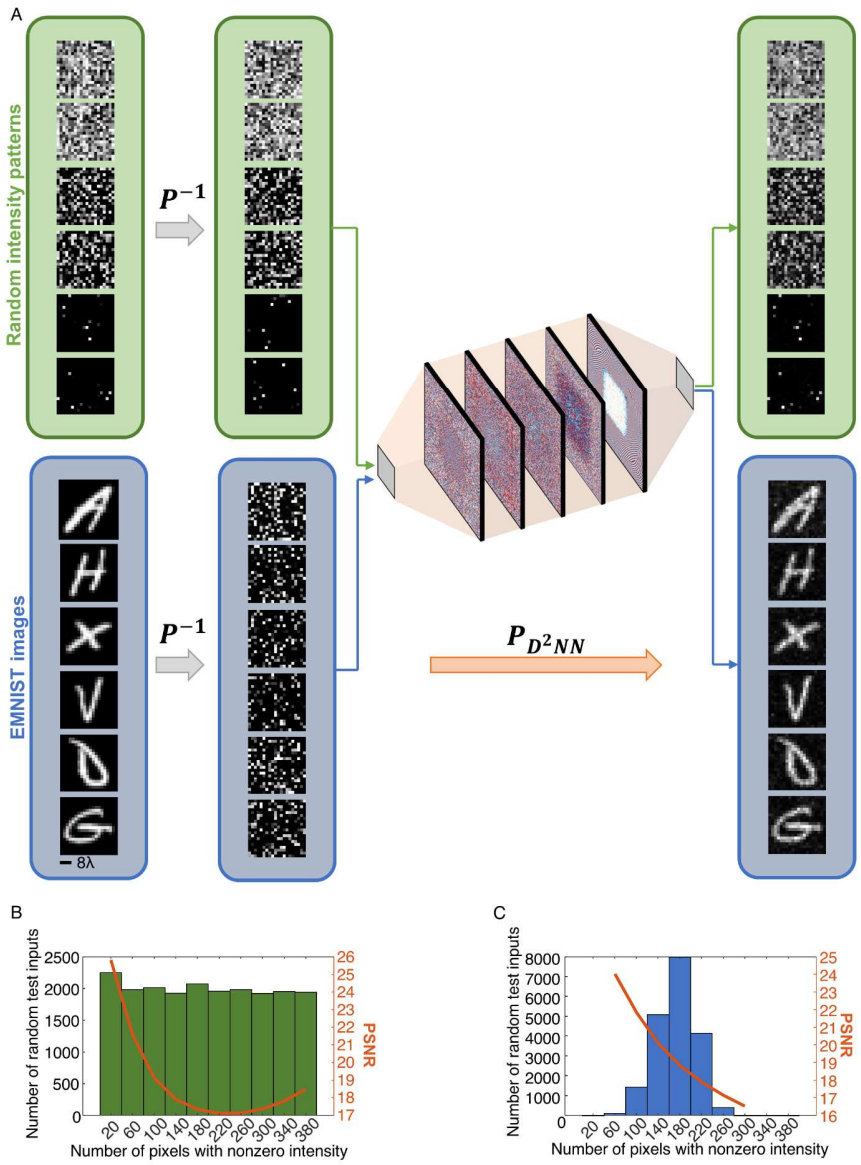

Fig. 6.2 Input-output intensity pairs. A The diffractive permutation network shown in Fig. 6.1 was tested on two different datasets. The first blind testing dataset contains 20K randomly generated inputs. 6 examples from this randomly created testing data are shown here for demonstrating input-output intensity pairs with low, moderate and high signal sparsity levels. Beyond successfully permuting randomly generated intensity patterns, the performance of the diffractive permutation network was also quantified using permuted EMNIST images. None of these test samples were used in the training phase. B Output intensity image PSNR with respect to the ground truth intensity patterns as a function of the input signal sparsity in randomly generated test dataset. C Same as B, except for EMNIST test images.

SSIM metrics, respectively, demonstrating the generalization capability of the diffractive network to new types of input image data never seen during the training phase.

Impact of number of diffractive layers and features

Next, we investigate the performance of diffractive permutation networks as a function of the number of diffractive neurons trained within the system. Towards this goal, in addition to the 5-layer design shown in Fig. 6.1 that has in total of N=200K trainable diffractive features, we trained diffractive permutation networks consisting of 4, 3 and 2 modulation surfaces. The physical design parameters such as the size/width of the diffractive surfaces, layer-to-layer distances and the extent of the input and output fields-of-view, were kept identical to the ones in the 5-layer network design. In other words, these new diffractive networks are designed and trained exactly in the same way as the previous 5-layer network except they contain fewer diffractive layers. Figures 6.3A and 6.3B provide a quantitative comparison between these 4 diffractive permutation networks. While Fig. 6.3A illustrates the mean PSNR and SSIM values achieved by each diffractive network for recovering EMNIST images, Fig. 6.3B demonstrates the mean-squared-error (MSE) between the desired permutation operation and its optically realized version (P_{D^2NN}) as a function of the number of diffractive layers utilized in these designs. According to the permutation operator error shown in Fig. 6.3B, the performance improvement of the system increases drastically with the additional diffractive layers up to the 4-layer design that represents a critical point in the sense that the inclusion of a 5th diffractive surface brings a relatively small improvement. The reason behind this behavior is the fact that the number of diffractive features, N, in the 4-layer diffractive permutation network matches the space-bandwidth product set by our input and output FOVs, i.e., $N_i N_o = 400\times400$=160K. In

other words, Fig. 6.3B reveals that when the number of phase-only diffractive modulation units N matches or exceeds $N_i N_o$, the diffractive optical network can achieve a given linear transformation between the input and output intensities with a very low error, i.e., $P_{D^2NN} \approx P$; for example, the MSE between P_{D^2NN} and P in the case of a 4-layer design was found to be 6.63×10^{-5}. For $N < N_i N_o$, the error between P_{D^2NN} and P increases accordingly, as shown in Fig. 6.3B.

The benefit of having $N \geq N_i N_o$ is further revealed in the increased generalization capability of the diffractive network as shown in Fig. 6.3A. Since the EMNIST images were not used during the training, they represent completely new types of input intensity patterns for the presented diffractive optical networks. The SSIM (PSNR) values achieved by the 4-layer diffractive network is found as 0.75 (16.41 dB) for the optical recovery of the permuted EMNIST images. These numbers are significantly higher compared to the performance of the 3-layer and 2-layer diffractive designs that can attain SSIM (PSNR) values of 0.46 (12.91 dB) and 0.30 (12.08 dB) for the same task; furthermore, the 5-layer diffractive network design shown in Fig. 6.1 outperforms the others by achieving 0.85 (19.18 dB) for the same performance metrics. The visual comparison of the input-output intensity patterns depicted in Fig. 6.3C further supports this conclusion, where the noise due to the crosstalk between interconnection channels decreases proportional to the number of diffractive layers in the system.

Vaccination of diffractive permutation networks

With sufficiently large number of phase-only diffractive neurons/features, the diffractive networks can optically realize permutation operations with e.g., 0.16 million channels between the input and output pixels as shown in Fig. 6.3. In fact, the number of interconnects that can be

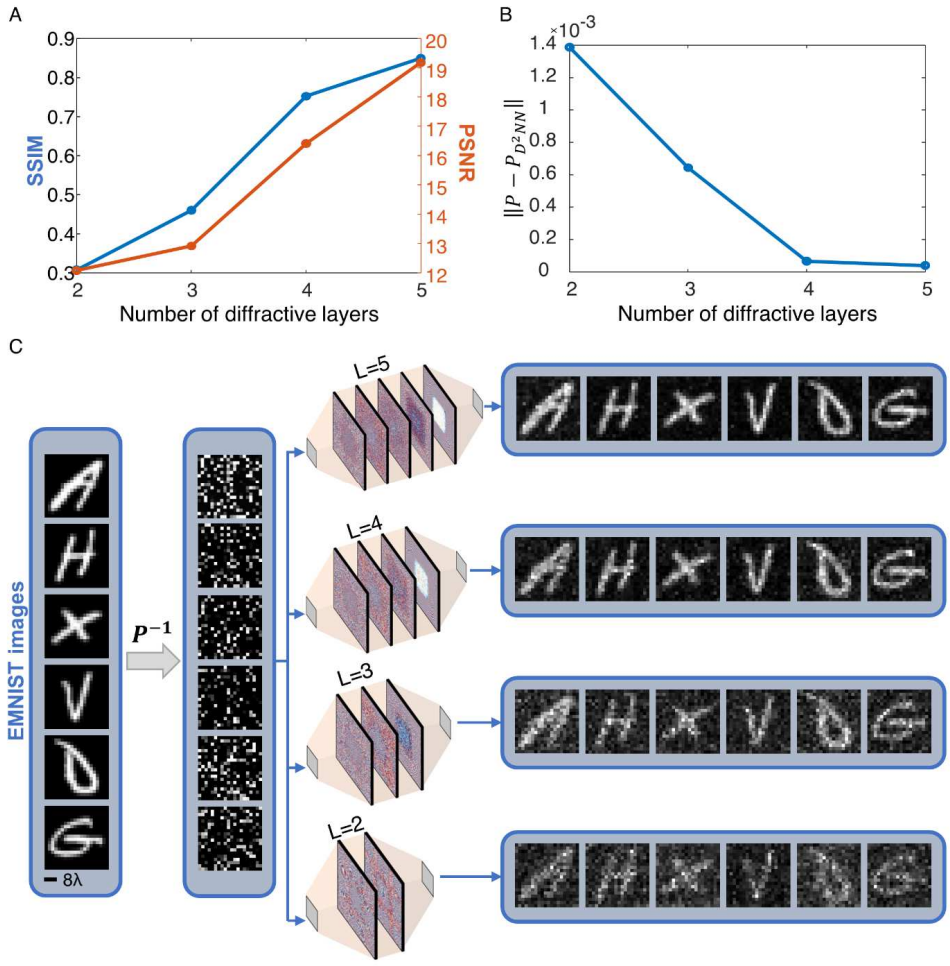

Fig. 6.3 The impact of the number of diffractive layers on the approximation accuracy of D2NN for a given intensity permutation operation. A The average SSIM and PSNR values achieved by the diffractive permutation network designs based on L=2, L=3, L=4 and L=5 diffractive layers, containing 200×200, i.e., 40K, phase-only diffractive neurons/features per layer for the task of optically recovering permuted EMNIST images. B The transformation error between the desired intensity permutation (P) and its optically realized counterpart ($P_{(D^2NN)}$) for the diffractive networks with L=2, L=3, L=4 and L=5 diffractive layers. The transformation error decreases as a function of the number of layers in the diffractive network architecture. The L=4-layer diffractive permutation network design represents a critical point as it matches the space-bandwidth product requirement of the desired permutation operation, i.e., N=Ni No=4×40K=160K, and further increasing the number of layers to L=5 brings only a minor improvement. C Examples of EMNIST test images demonstrating the performance of the diffractive permutation networks as a function of L.

optically implemented through diffractive networks can go far beyond 0.16 million, given that the size/width of the diffractive surfaces and the number of diffractive layers can be increased further depending on the fabrication technology and the optomechanical constraints of the system. In addition, as the number of diffractive layers increases in a diffractive network architecture, their forward model can better generalize to new, unseen data as shown in Fig. 6.3A.

On the other hand, deeper diffractive optical network designs are more susceptible to misalignments that are caused by the limitations of the optomechanical assembly and/or the fabrication technology that is utilized. It was shown that diffractive optical networks trained for statistical inference tasks e.g., all-optical object classification, can be vaccinated against misalignments and other physical error sources, when the factors creating these nonideal conditions were incorporated into the training forward model, which was termed as vaccinated-D^2NNs or v-D^2NNs[103]. Specifically, v-D^2NN expands on the original D^2NN framework by modeling possible error sources as random variables and integrating them as part of the training model so that the deep learning-based evolution of the diffractive surfaces is guided towards solutions that are resilient to nonideal physical conditions and/or fabrication errors. Towards practical applications of diffractive permutation networks, we quantified the impact of optomechanical errors and applied the v-D^2NN framework to devise robust solutions that can achieve a given interconnect operation despite fabrication tolerances.

In our numerical study depicted in Fig. 6.4, we considered 4 different misalignment components representing the 3D misalignment vector of the l^{th} diffractive layer, $\left({D_x}^l, {D_y}^l, {D_z}^l\right)$ and their in-plane rotation around the optical axis denoted as ${D_\theta}^l$. Each of these 4 misalignment components were defined as independent, uniformly distributed random variables,

${D_*}^l \sim U(-\Delta_*, \Delta_*)$, with Δ_* defined as a function of a common auxiliary parameter, v. The lateral misalignments parameters, Δ_x and Δ_y, determining the range of ${D_x}^l$ and ${D_y}^l$, respectively, were set to be $0.67\lambda v$, i.e., ${D_x}^l \sim U(-0.67\lambda v, 0.67\lambda v)$ and ${D_y}^l \sim U(-0.67\lambda v, 0.67\lambda v)$, where λ denotes the wavelength of the illumination light. Similarly, Δ_z and Δ_θ were defined as $24\lambda v$ and $4^\circ v$. For instance, if we take $v = 0.5$, this means each diffractive layer can independently/randomly shift in both x and y axes within a range of $(-0.335\lambda, 0.335\lambda)$. In addition, their location over the z direction and their in-plane orientation can randomly change within the ranges of $(-12\lambda, 12\lambda)$ and $(-2^\circ, 2^\circ)$, respectively (see the Methods section for more details).

To better highlight the impact of these misalignments and demonstrate the efficacy of the v-D^{2}NN framework, we trained a new nonvaccinated, i.e., $v_{tr} = 0$, diffractive permutation network that can all-optically realize a given permutation matrix, $\boldsymbol{P}$, representing 10K intensity interconnections between 100 input and 100 output pixels of size 4λ×4λ. The error-free training model of this diffractive network with $v_{tr} = 0$ implicitly assumes that when the resulting diffractive network is fabricated, the system conditions will exactly match the ideal settings regarding the 3D locations of the layers and their in-plane orientations. With an architecture identical to the one shown in Fig. 6.1, containing $N = 200K \gg N_i N_o$ diffractive neurons, this diffractive network can all-optically approximate the permutation matrix, $\boldsymbol{P}$, with an MSE of 1.45×10^{-6} in the absence of any misalignment errors, i.e., $v_{test} = 0$ (see the green curve in Fig. 6.4B). However, when there is some discrepancy between the training and testing conditions, i.e., $v_{test} > 0$, the optically implemented forward transformation, $\boldsymbol{P_{D^2NN}}$, starts to deviate from the desired operation $\boldsymbol{P}$. For instance, at $v_{test} = 0.125$, the transformation error, $\left\| \boldsymbol{P_{D^2NN}} - \boldsymbol{P} \right\|$, can

be computed as 3.1×10^{-3}. This negative impact of the physical misalignments on the performance of a *nonvaccinated* diffractive network can also be seen in Fig. 6.4A (green curve), which demonstrates the SSIM values achieved by this diffractive network for recovering permuted EMNIST images under different levels of misalignments. The high-quality of the image recovery (see Fig. 6.4C) at $v_{test} = 0$ quantified with an SSIM of 0.99 deteriorates under the presence of misalignments, highlighted by the SSIM value falling to 0.49 and 0.30 at $v_{test} = 0.125$ and $v_{test} = 0.25$, respectively.

Unlike the nonvaccinated design, the vaccinated diffractive permutation networks can maintain their approximation capacity and accuracy under erroneous testing conditions as shown in Figs. 6.4A-B. For instance, the SSIM value of 0.49 attained by the nonvaccinated diffractive network for the misalignment uncertainty set by $v_{test} = 0.125$, increases to 0.88 in the case of a diffractive permutation network trained with $v_{tr} = 0.25$ (red curve in Fig. 6.4A). The difference between the image recovery performances of the vaccinated and the nonvaccinated diffractive network designs increases further as the misalignment levels increase during the blind testing. While the nonvaccinated diffractive network can only achieve SSIM values of 0.3 and 0.19 at $v_{test} = 0.25$ and $v_{test} = 0.375$, respectively, the output images synthesized by the vaccinated design ($v_{tr} = 0.25$) reveals SSIM values of 0.8 at $v_{test} = 0.25$ and 0.64 at $v_{test} = 0.375$ (see Fig. 6.4D). A similar conclusion can also be drawn from Fig. 6.4B, demonstrating the MSE values between the desired permutation matrix, $\boldsymbol{P}$, and its optically realized counterpart, $\boldsymbol{P_{D^2NN}}$. The transformation errors, $\left\|\boldsymbol{P_{D^2NN}} - \boldsymbol{P}\right\|$, of the vaccinated diffractive network ($v_{tr} = 0.25$) at $v_{test} = 0.125$ and at $v_{test} = 0.25$ were computed as 5.15×10^{-4} and 1.2×10^{-3}, respectively,

which are 5-10 times smaller compared to the MSE values provided by the nonvaccinated diffractive design at the same misalignment levels.

The compromise for this misalignment robustness comes in the form of a reduction in the peak performance. While the nonvaccinated diffractive network can solely focus on realizing the given permutation operation with the highest quality and approximation accuracy, the vaccinated diffractive network designs partially allocate their degrees-of-freedom to building up resilience against physical misalignments. For example, while the peak SSIM achieved by the nonvaccinated diffractive network is 0.99, it is 0.88 for the diffractive permutation network vaccinated with $v_{tr} = 0.25$. The key difference, on the other hand, is that the better performance of the nonvaccinated diffractive network is sensitive to the physical implementation errors, while the vaccinated diffractive permutation networks can realize the desired input-output interconnects over a larger range of fabrication errors or tolerances. A comparison between the diffractive layer patterns of the nonvaccinated and vaccinated diffractive permutation networks shown in Figs. 6.4C and 6.4D, respectively, also reveals that the vaccination strategy results in smoother light modulation patterns; in other words, the material thickness values over the neighboring diffractive neurons partially lose their independence and become correlated, causing a reduction in the number of independent degrees-of-freedom in the system.

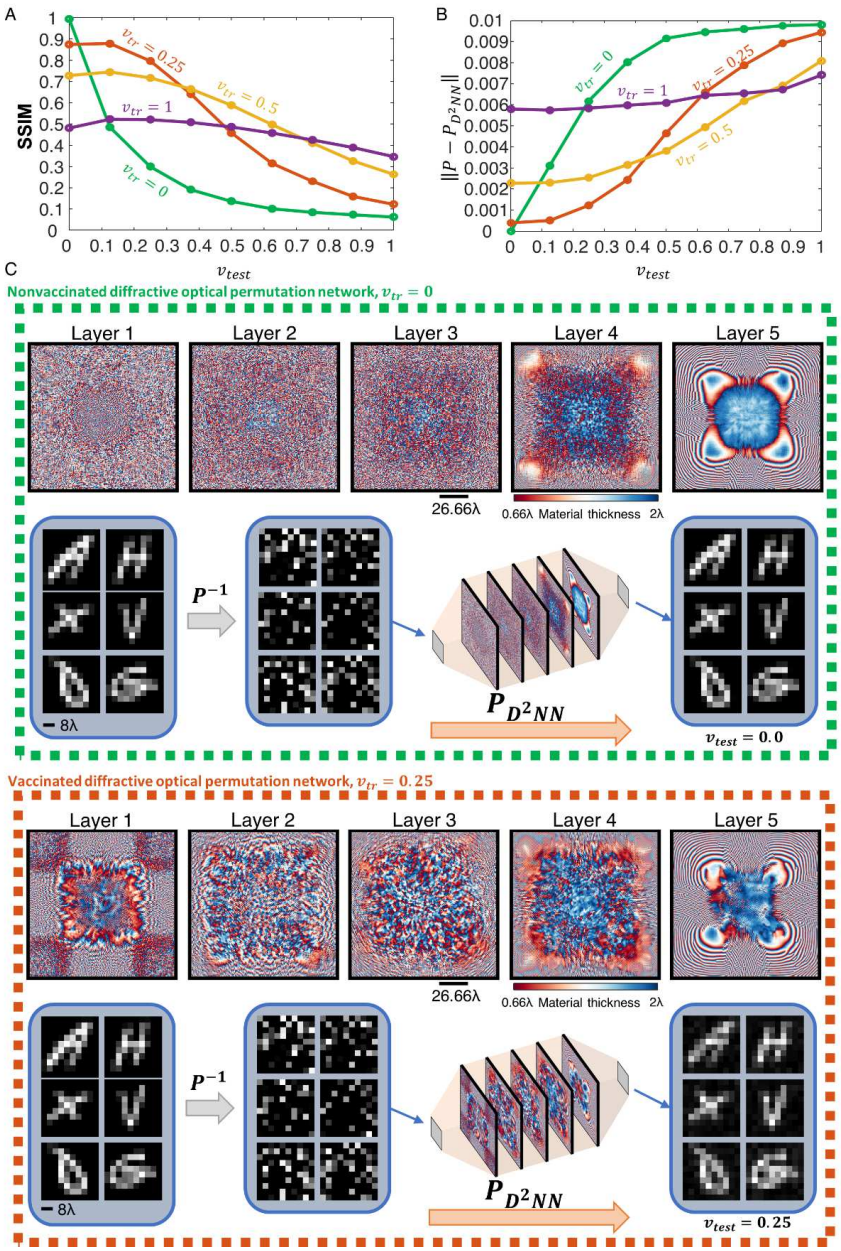

Fig. 6.4 The sensitivity of the diffractive permutation networks against various levels of physical misalignments. A SSIM values achieved by 5-layer diffractive permutation networks with and without vaccination. B Transformation errors between the desired 100×100 permutation operation (P) and its optically synthesized counterpart ($P_{(D^2 NN)}$) at different levels of misalignments denoted by v_{test}. C The layers of a nonvaccinated diffractive permutation network, i.e., v_{tr}=0, along with the examples of EMNIST test images recovered optically through the diffractive permutation operation. D Same as C, except for a vaccinated diffractive permutation network based on v_{tr}=0.25.

Experimental demonstration of a diffractive permutation network

To experimentally demonstrate the success of the presented diffractive permutation interconnects, we designed a 3-layer diffractive permutation network achieving the desired (randomly generated) intensity shuffling operation with $N_i = N_o = 5x5$, optically synthesizing 625 connections between the input and output FOVs; this network was designed to operate at 0.4 THz, corresponding to ~0.75 mm in wavelength. During the training, the forward model of this diffractive permutation network was vaccinated with $v_{tr} = 0.5$ against the 4 error sources as detailed in Section 2.2 including the 3D location of each diffractive layer and the in-plane rotation angle around the optical axis. In addition to these misalignment components, we also vaccinated this diffractive network model against unwanted material thickness variations that could arise due to the limited lateral and axial resolution of our 3D printer (see the Methods section for details). To compensate for the reduction in the degrees-of-freedom due to the vaccination scheme, the number of phase-only diffractive features in the permutation network was selected to be $N_L = 10$K diffractive neurons per layer. Therefore, each diffractive layer shown in Fig. 6.5A contains 100×100 phase-only diffractive neurons of size ~0.67λ×0.67λ. Compared to the diffractive surfaces shown in Figs. 6.1-6.4, the layers of our experimental system were set to be 2-times smaller in both the x and y directions to keep the layer-to-layer distances smaller while maintaining the level of optical connectivity between the successive diffractive surfaces (see Fig. 6.5B). Figures 6.5C and 6.5D illustrate the 3D printed diffractive permutation network and the schematic of our experimental setup (see the Methods section for details).

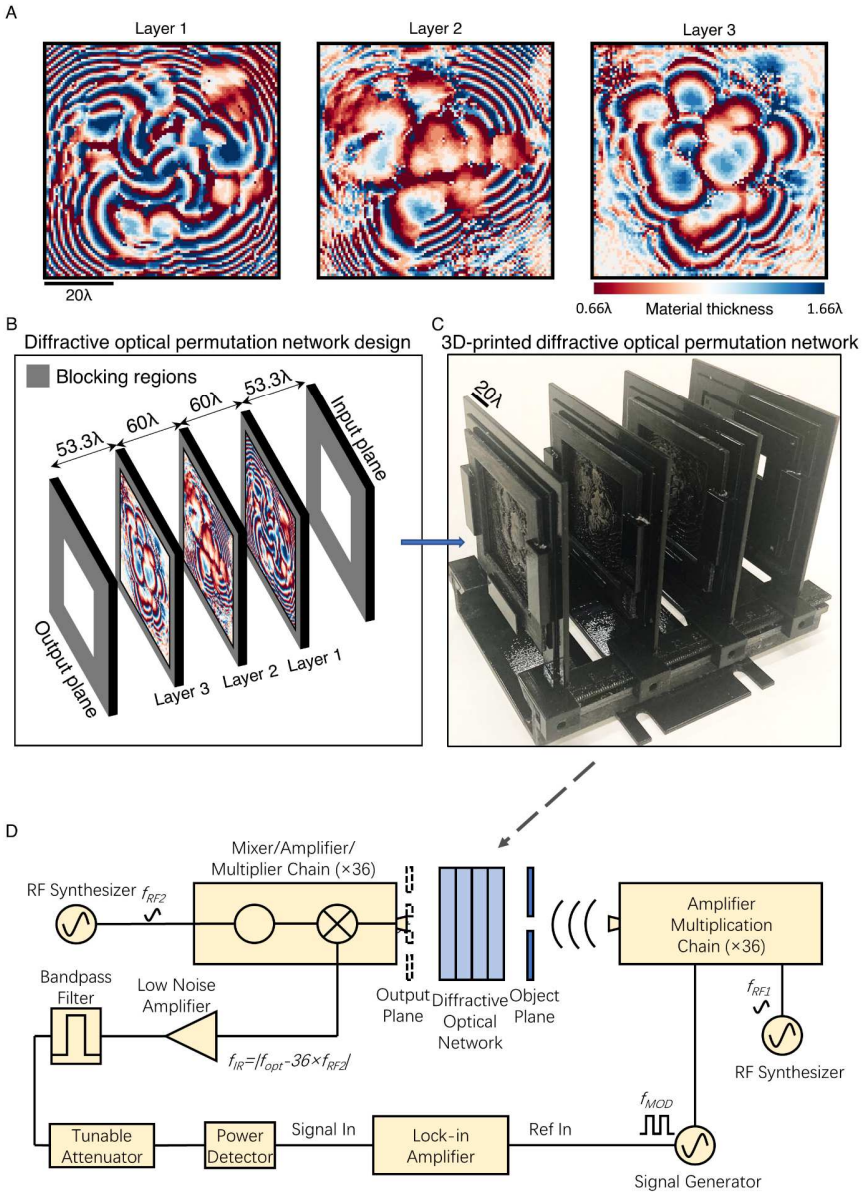

Fig. 6.5 Experimental demonstration of a diffractive permutation network. A The material thickness profiles of the diffractive surfaces of the fabricated diffractive permutation network. B The schematic of the experimental architecture illustrating the forward optical model of the diffractive permutation network. C 3D printed diffractive permutation network operating at THz part of the spectrum. D The schematic of our experimental system.

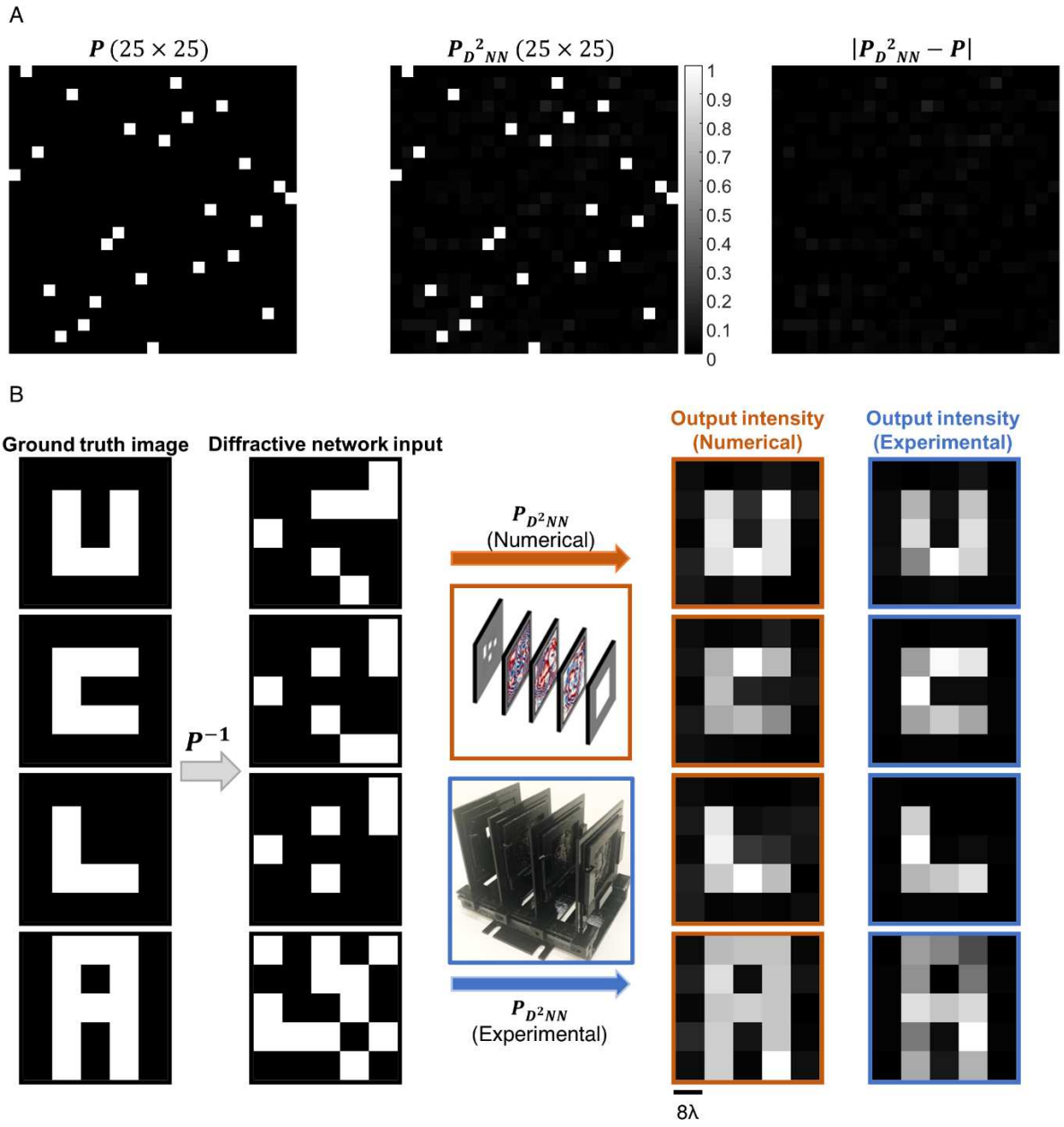

Fig. 6.6 Experimental results. A (left) The desired 25×25 permutation matrix, P, (middle) the optically realized permutation operation predicted by the numerical forward model, P_(D^2 NN), and (right) the absolute error map between the two matrices. B Comparison between the numerically predicted and the experimentally measured output images for the task of recovering intensity patterns describing the letters 'U', 'C', 'L' and 'A'.

Figure 6.6A illustrates the targeted 25×25 permutation matrix ($\boldsymbol{P}$) that is randomly generated and the numerically predicted $\boldsymbol{P}_{\boldsymbol{D^2NN}}$ along with the absolute difference map between these two matrices. According to the numerical forward model of the trained diffractive network shown in Fig. 6.5, the transformation error between the $\boldsymbol{P}$ and $\boldsymbol{P}_{\boldsymbol{D^2NN}}$, i.e., $\left\|\boldsymbol{P}_{\boldsymbol{D^2NN}} - \boldsymbol{P}\right\|$ is equal to 5.99×10^{-4} under error-free conditions, i.e., $v_{test} = 0$. Furthermore, the forward model of the trained diffractive permutation network shown in Fig. 6.5 provides 17.87 dB PSNR on average for the test letters 'U', 'C', 'L' and 'A', as depicted in Fig. 6.6B. A visual comparison between the numerically predicted and the experimentally measured output images of these 4 input letters (which were never seen by the network before) demonstrates the accuracy of the forward training and testing models as well as the success of the presented diffractive permutation network design. Interestingly, the PSNR of the experimentally measured images was observed to be higher, 19.54 dB, compared to the numerically predicted value, 17.87 dB. Our numerical study reported in Fig. 6.4 suggests that this can be explained based on the vaccination range used during the training and the amount physical error in the system testing. For instance, the SSIM value achieved by the vaccinated diffractive network trained with $v_{tr} = 0.5$ (yellow curve) at relatively lower physical misalignment levels, e.g., $v_{test} = 0.125$, is higher compared to its performance under the ideal conditions, i.e., $v_{test} = 0.0$, as depicted in Fig. 6.4A.

6.3 Discussion

Beyond optomechanical error sources and fabrication tolerances, another factor that might potentially hinder the utilization of diffractive permutation networks in practical applications is the output diffraction efficiency. For instance, the diffraction efficiency of the 5-layer network shown in Fig. 6.1 is ~0.004% which might be very low for some applications. On the other hand, this can be significantly increased by using an additional loss term, penalizing the poor

diffraction efficiency of the network (see the Methods section for details). The training of these diffractive network models is based on a loss function in the form of a linear combination of two different penalty terms, $\mathcal{L}' = \mathcal{L} + \gamma\mathcal{L}_e$, where $\mathcal{L}$ is a structural loss term enforcing transformation quality/accuracy and $\mathcal{L}_e$ is the diffraction efficiency related penalty term promoting efficient solutions (see the Methods section). As a general trend, the diffraction efficiency of the underlying diffractive network model increases as a function of the weight (γ) of the efficiency penalty term in the loss function. However, since the number of diffractive neurons, hence, the degrees-of-freedom in these diffractive network models is very close to $N_i N_o$, the diffraction efficiency either does not improve beyond a certain value or the evolution of the diffractive layers starts to solely focus on the efficiency instead of the desired permutation operation resulting in low performance designs. This unstable behavior can be observed specifically when $0.235 < \gamma < 0.24$. On the other hand, as in the case of vaccinated diffractive network models, if the diffractive network architecture contains $N \gg N_i N_o$ diffractive neurons, then this instability vanishes, providing significant improvements in the output diffraction efficiency without sacrificing the performance of the all-optical permutation operation. For instance, the 3D printed diffractive permutation network depicted in Fig. 6.6 was trained based on $\mathcal{L}'$ with γ = 0.15 and it provides 2.45% output diffraction efficiency, despite the fact that 89.37% of the incident power at the input plane is lost due to the absorption of the 3D printing material. With weakly absorbing transparent materials used as part of the diffractive network fabrication, a significantly larger output efficiency can be achieved.

Also note that, although we solely focused on diffractive network designs composed of dielectric optical modulation surfaces, in principle, some of these layers can be replaced with metasurfaces/metamaterials. While the use of metamaterials can provide some additional degrees

of freedom, including, for example, the engineering of dispersion, there are some challenges to overcome in realizing metamaterial-based diffractive networks. First, metamaterials could lead to crosstalk between the secondary wave fields created by adjacent meta-atoms. Although the lateral pitch between neighboring meta-units can be increased to partially mitigate this field crosstalk problem, such an approach would sacrifice the lateral density of meta-units packed over each diffractive surface leading to reduced computational capacity. Moreover, in the presence of fabrication errors and imperfections, the scattered light fields might deviate from the predictions of the numerical forward model. These nonideal waves generated by a metasurface would then excite unwanted diffraction modes over the subsequent layers generating an "avalanche" within the diffractive volume, accumulating substantial field errors, especially for deeper network designs with $L \geq 2$. In addition, the physical models of phase and/or amplitude modulation of meta-atoms are, in general, valid for waves covering a relatively small numerical aperture (NA). As a result, a high NA diffractive network design that utilizes the entire bandwidth of the propagation medium (NA=1, for air), would be challenging, as the modulation response of the meta-units might deviate from their ideal small angle responses, introducing errors to the forward model. Although it is possible to restrict a diffractive network design to work with a lower NA, it would increase the overall footprint of the system and reduce the space-bandwidth product that can be processed by the diffractive network.

These challenges, in general, are negligible for dielectric diffractive networks composed of $\lambda/2$ features on a substrate, as also highlighted by the close match between the numerically predicted images and their experimentally measured counterparts shown in Fig. 6.6 and our former work[70,73,75,77-78]. In our optical forward model, the diffractive layers are assumed to be thin modulation surfaces, i.e., there is only a single scattering event converting an incident wave

field to an outgoing one after each diffractive layer. Practically, though, there are additional scattered fields that are ignored in our model, especially when there is a substantial material thickness variation between adjacent pixels. However, in our designs, we do not observe a sharp material thickness transition between neighboring pixels. This is mainly due to the nature of our training process. Specifically, the presented diffractive networks are trained through error-backpropagation, which computes the variable updates by taking the gradient of the loss function with respect to the material thickness over each diffractive unit. In such a process, it is highly unlikely that the gradients of two adjacent pixels ($\lambda/2$ apart from each other) deviate significantly from each other, which effectively causes a smoothening effect on the diffractive surface height profiles as they are being optimized through deep learning. This smoothening behavior is even more pronounced in the vaccinated diffractive network designs due to the random lateral translation of the diffractive layers as part of the training forward model. Therefore, the impact of side scattering or field shadowing due to height discontinuities across a given diffractive layer design is negligible. In addition, the back-reflected waves can also be ignored, as these are, in general, weak processes unless they are specifically enhanced using, e.g., metamaterials or other special structures. Therefore, the optical forward model of dielectric diffractive networks can be accurately represented within the scalar diffraction theory without needing vectorial modeling of light fields or considering weaker multi-reflections. Finally, the evanescent waves and the vectorial fields associated with them can be entirely ignored since each successive diffractive layer is axially positioned $>\lambda$ away from the previous layer.

In summary, we showed that the diffractive networks can optically implement intensity permutation operations between their input and output apertures based on phase-only light modulation surfaces with $N \geq N_i N_o$ diffractive neurons. Due to the nonlinear nature of the

intensity operation, it is crucial to use training input intensity patterns with different levels of sparsity to prevent any type of data-specific overfitting during the training phase. Diffractive permutation networks with $N > N_i N_o$ demonstrate increased generalization capability, synthesizing more accurate outputs with $\left\| \boldsymbol{P}_{\boldsymbol{D^2NN}} - \boldsymbol{P} \right\| \approx 0$. By using $N > N_i N_o$ one can also design misalignment and fabrication error insensitive, power-efficient diffractive permutation networks, which could play a major role in practical applications, e.g., 6G wireless networks, computational cameras, etc. Although this study demonstrated diffractive optical networks realizing permutation operations with 0.16 million interconnects, with $N_i = N_o = 20 \times 20$, these systems are highly scalable to even larger N_i, N_o combinations depending on the availability of training computer hardware. Since the training of a diffractive optical network is a one-time effort, one can use a computing platform with a significantly larger random-access memory (RAM) to design much bigger diffractive networks. Alternatively, the forward training model of a diffractive network can also be distributed among multiple GPUs for parallel computing with increased memory capacity paving the way to significantly larger permutation operations to be implemented all-optically. Finally, the incorporation of dynamic spatial light modulators to replace some of the diffractive layers in a given design can be used to reconfigure, on demand, the all-optically performed diffractive transformation.

6.4 Materials and Methods

Experimental setup

According to the schematic diagram of our experimental setup shown in Fig. 6.5D, the THz wave incident on the input FOV of the diffractive network was generated using a horn antenna attached to the source WR2.2 modulator amplifier/multiplier chain (AMC) from

Virginia Diode Inc. (VDI). A 10 dBm RF input signal at 11.111 GHz (f_{RF1}) at the input of the AMC was multiplied 36 times to generate a continuous-wave (CW) radiation at 0.4 THz, corresponding to ~0.75 mm in wavelength. The output of the AMC was modulated with 1 kHz square wave to resolve low-noise output data through lock-in detection. Since we did not use any collimating optics in our setup, the distance between the input plane of the 3D-printed diffractive optical network and the exit aperture of the horn antenna was set to be ~60 cm approximating a uniform plane wave over the 40λ × 40λ input FOV. At the output plane of the diffractive optical network, the diffracted THz light was collected using a single-pixel Mixer/AMC from Virginia Diode Inc. (VDI). During the measurements, the detector received a 10 dBm sinusoidal signal at 11.083 GHz serving as a local oscillator for mixing, and the down-converted signal was at 1GHz. The 40λ × 40λ output FOV was scanned by placing the single-pixel detector on an XY stage that was built by combining two linear motorized stages (Thorlabs NRT100). At each scan location, the down-converted signal coming from the single-pixel detector was fed to low-noise amplifiers (Mini-Circuits ZRL-1150-LN+) with a gain of 80 dBm and a 1 GHz (+/-10 MHz) bandpass filter (KL Electronics 3C40-1000/T10-O/O) that erases the noise components coming from unwanted frequency bands. Following the amplification and filtering, the measured signal passed through a tunable attenuator (HP 8495B) and a low-noise power detector (Mini-Circuits ZX47-60). Finally, the output voltage value was generated by a lock-in amplifier (Stanford Research SR830). The modulation signal was used as the reference signal for the lock-in amplifier and accordingly, we performed a calibration to convert the lock-in amplifier readings at each scan location to linear scale. During our experiments, the scanning step size at the output plane was set to be ~λ in x and y directions. The smallest pixel of the experimentally targeted permutation grid, i.e., the desired resolution of the diffractive permutation operation was taken as 8λ × 8λ during the

training, corresponding to 5 × 5 discrete input and output signals. Therefore, the output signal measured for each input object was integrated over a region of 8λ × 8λ per pixel, resulting in the measured images shown in Fig. 6.6B.

A 3D printer, Objet30 Pro, from Stratasys Ltd., was used to fabricate the layers of the diffractive permutation network shown in Fig. 6.5C as well as the layer holders. The active modulation area of our 3D printed diffractive layers was 5 cm × 5 cm (~66.66λ × ~66.66λ) containing 100 × 100, i.e., 10K, diffractive neurons. These modulation surfaces were printed as insets surrounded by a uniform slab of printing material with a thickness of 2.5 mm and the total size of each printed layer including these uniform regions was 6.2 cm × 6.2 cm. Following the 3D printing, these additional surrounding regions were coated with aluminum to block the propagation of the light over these areas minimizing the contamination of the output signal with unwanted scattered light.

Optical forward model of diffractive permutation networks

The material thickness, h, was selected as the physical parameter controlling the complex-valued transmittance values of the diffractive layers of our design. Based on the complex-valued refractive index of the diffractive material, $\tau = n + j\kappa$, the corresponding transmission coefficient of a diffractive neuron located on the l^{th} layer at a coordinate of (x_q, y_q, z_l) is defined as,

$$t(x_q, y_q, z_l) = \exp\left(\frac{-2\pi\kappa h(x_q, y_q, z_l)}{\lambda}\right)\exp\left(\frac{-j2\pi(n - n_m)h(x_q, y_q, z_l)}{\lambda}\right) \quad (6.1)$$

where $n_m = 1$ denotes the refractive index of the propagation medium (air) between the layers. The real and imaginary parts of the 3D printing material were measured experimentally using a THz spectroscopy system and they were revealed as $n = 1.7227$ and $\kappa = 0.031$ at 0.4 THz.

The optical forward model of the presented diffractive networks relies on the Rayleigh-Sommerfeld theory of scalar diffraction to represent the propagation of light waves between the successive layers. According to this diffraction formulation, the free-space can be interpreted as a linear, shift-invariant operator with the impulse response,

$$w(x, y, z) = \frac{z}{r^2}\left(\frac{1}{2\pi r} + \frac{n}{j\lambda}\right)\exp(\frac{j2\pi nr}{\lambda}) \tag{6.2}$$

where $r = \sqrt{x^2 + y^2 + z^2}$. Based on Eq. 6.2, q^{th} diffractive neuron on the l^{th} layer, at (x_q, y_q, z_l), can be interpreted as the source of a secondary wave generating the field at (x, y, z) in the form of,

$$w_q^l(x, y, z) = \frac{z - z_l}{\left(r_q^l\right)^2}\left(\frac{1}{2\pi r_q^l} + \frac{n}{j\lambda}\right)\exp\left(\frac{j2\pi nr_q^l}{\lambda}\right). \tag{6.3}$$

The parameter r_q^l in Eq. 6.3 is expressed as $\sqrt{\left(x - x_q\right)^2 + \left(y - y_q\right)^2 + (z - z_l)^2}$. When each diffractive neuron on layer l generates the field described by Eq. 6.3, the light field incident on the p^{th} diffractive neuron on the $(l+1)^{th}$ layer at (x_p, y_p, z_{l+1}) is the linear superposition of the all the secondary waves generated by the previous layer l, i.e., $\sum_q A_q^l w_q^l\left(x_p, y_p, z_{l+1}\right)$, where A_q^l is the complex amplitude of the wave field right after the q^{th} neuron of the l^{th} layer. This field is modulated by the multiplicative complex-valued transmittance of the diffractive unit at

(x_p, y_p, z_{l+1}), creating the modulated field $t(x_p, y_p, z_{l+1}) \sum_q A_q^l w_q^l(x_p, y_p, z_{l+1})$. Based on this new modulated field, a new secondary wave,

$$u_p^{l+1}(x, y, z) = w_p^{l+1}(x, y, z) t(x_p, y_p, z_{l+1}) \sum_q A_q^l w_q^l(x_p, y_p, z_{l+1}), \quad (6.4)$$

is generated. The outlined successive modulation and secondary wave generation processes occur until the waves propagating through the diffractive network reach to the output plane. Although, the forward optical model described by Eqs. 6.1-6.4 is given over a continuous 3D coordinate system, during our deep learning-based training of the presented diffractive permutation networks, all the wave fields and the modulation surfaces were represented based on their discrete counterparts with a spatial sampling rate of ~0.67λ on both x and y axes, that is also equal to the size of a diffractive neuron.

Physical architecture of diffractive permutation networks

The size of the output and input FOVs of the presented diffractive permutation networks were both set to be 40λ × 40λ, defining a unit magnification optical permutation operation. Note that the unit magnification is not a necessary condition for the success of the forward operation of diffractive optical interconnects but rather a design choice. Without loss of generality, the output FOV can be defined centered around the origin, $(0,0)$, i.e., $-20\lambda < x, y < 20\lambda$. The dimensions of the diffractive layers was taken as 133.3λ×133.3λ for the diffractive permutation networks presented in Figs. 6.1-6.4, and in all these diffractive network architectures the layer-to-layer distances were taken as 120λ. The axial distance between the 1st diffractive layer and the input FOV was set to be 53.3λ that is also equal to the axial distance from the last diffractive layer to the output plane, preserving the symmetry of the system on the longitudinal axis. In the

case of our experimentally validated diffractive design (Fig. 6.5), on the other hand, the active modulation surface of the fabricated diffractive layers extends 66.7λ on both x and y directions. Accordingly, the layer-to-layer distances were taken as 60λ while the remaining distances were kept equal to 53.3λ.

During the deep learning-based training of all of these diffractive permutation networks, the wave fields and the propagation functions depicted in Eqs. 6.2-6.4 were sampled at a rate of ~0.67λ that is also equal to the size of the smallest diffractive units on the modulation surfaces constituting the presented diffractive networks. At this spatial sampling rate, the input and output intensity patterns were represented as 2D discrete vectors of size 60×60 denoted by $I_{in}[m,n]$ and $I_{out}[m,n]$, respectively, with $m = 1,2,3,\dots,60$ and $n = 1,2,3,\dots,60$. The underlying complex-valued wave fields can be written as $U_{in}[m,n] = \sqrt{I_{in}[m,n]}e^{j\emptyset_{in}[m,n]}$ and $U_{out}[m,n] = \sqrt{I_{out}[m,n]}e^{j\emptyset_{out}[m,n]}$. In our forward model, we assumed that the input light has constant phase front, i.e., $\emptyset_{in}[m,n]$ is taken as an arbitrary constant within the input field-of-view. In alternative implementations, without loss of generality, the diffractive permutation network can be trained with any arbitrary function of $\emptyset_{in}[m,n]$, achieving the same output accuracy levels $\left\|\boldsymbol{P}_{\boldsymbol{D^2NN}} - \boldsymbol{P}\right\| \approx 0$ using $N \geq N_i N_o$.

While the light fields, the diffractive layers and the impulse response of the free-space were all sampled at a rate of ~0.67λ, the spatial grid/pixel size of a given desired permutation operation was set to be larger. Specifically, the permutation pixel size was taken as 2λ×2λ for the diffractive networks shown in Figs. 6.1-6.3. On the other hand, the input and output pixel size was chosen as 4λ×4λ for the vaccinated and nonvaccinated diffractive permutation networks

shown in Fig. 6.4; and finally, the pixel size was set to be 8λ×8λ for the fabricated diffractive permutation network model depicted in Fig. 6.5.

To train the presented diffractive permutation networks, a structural loss function, $\mathcal{L}$, in the form of MSE was used.

$$\mathcal{L} = \frac{1}{S}\sum_{s=1}^{S} |\boldsymbol{P} I_{in}[s] - \sigma I_{out}[s]|^{\mathbf{2}}, \tag{6.5}$$

In Eq. 6.5, $I_{in}[s]$ and $I_{out}[s]$ denote the lexicographically ordered vectorized counterparts of the input intensity pattern, i.e., $\mathrm{vec}(I_{in}[q,p])$, and the output intensity pattern, i.e., $vec(I_{out}[q,p])$, and $\boldsymbol{P}$ represents the desired permutation matrix to be performed all-optically. As depicted in Eq. 6.5, the output intensity pattern $I_{out}[s]$ or $I_{out}[q,p]$ was scaled by a constant σ that was calculated at each training iteration as,

$$\sigma = \frac{\frac{1}{S}\sum_{s=1}^{S} \boldsymbol{P} I_{in}[s] I_{out}[s]}{\frac{1}{S}\sum_{s=1}^{S} I_{out}[s]^{\mathbf{2}}}. \tag{6.6}$$

Note that the presented diffractive permutation networks preserve the relative intensity levels. Stated differently, our training forward model aims to keep the intensity levels over the output and input pixels the same up to a single multiplicative constant, σ.

To improve the diffraction efficiency of diffractive permutation networks, we defined another loss function, $\mathcal{L}'$, that is a linear combination of two penalty terms, $\mathcal{L}' = \mathcal{L} + \gamma\mathcal{L}_e$, where $\mathcal{L}$ corresponds to the structural loss defined in Eq. 6.5. $\mathcal{L}_e$ is the penalty term that promotes higher diffraction efficiency at the output of diffractive networks, and it was defined as, $\mathcal{L}_e = e^{-\eta}$, where,

$$\eta = \frac{\sum_{s=1}^{S} I_{out}[s]}{\sum_{s=1}^{S} I_{in}[s]} \times 100. \tag{6.7}$$

The diffractive permutation networks presented in Figs. 6.1-6.4 were trained based on $\mathcal{L}'$ with $\gamma = 0$; however, the experimentally demonstrated diffractive permutation network model was trained with $\gamma = 0.15$, resulting in an output diffraction efficiency of 2.45% (which includes a material absorption loss of 89.37%).

The supervised deep learning-based training of the presented diffractive permutation networks evaluates the loss function $\mathcal{L}'$ for a batch of randomly generated input patterns, computes the mean gradient and updates the learnable, auxiliary variables, h_a, that determine the material thickness over each diffractive neuron, h, through the following relation,

$$h(h_a) = \frac{\sin(h_a) + 1}{2}(h_m - h_b) + h_b \tag{6.8}$$

where h_m and h_b denote the maximum modulation thickness and the base material thickness, respectively. For all the diffractive permutation networks presented in Figs. 6.1-6.4 h_m was taken as 2λ. In the design of the 3D-printed diffractive permutation network, however, h_m was set to be 1.66λ to restrict the material thickness contrast between the neighboring diffractive features. The value of h_b was taken as 0.66λ for all the presented designs including the fabricated diffractive network.

Computation of $\boldsymbol{P_{D^2NN}}$, optical transformation errors and performance quality metrics

For a given diffractive permutation network design trained to optically implement a permutation matrix $\boldsymbol{P}$ of size $N_i \times N_o$, there are two different ways to compute the permutation operation predicted by its numerical forward model. The first way is to propagate N different

randomly generated independent inputs with $N \geq N_i N_o$ and solve a linear system of equations for revealing the entries of P_{D^2NN}. Alternatively, each input pixel at the input FOV can be turned on sequentially and the output intensity pattern synthesized by the diffractive optical permutation network as a response to each pixel provides one unique column of P_{D^2NN}. These two procedures, in general, result in two different P_{D^2NN} matrices that closely resemble each other. We opted to use the latter procedure due to its simplicity, which turn on each input pixel one at a time and records the corresponding output intensity pattern, which, after vectorization, represents a column of $\boldsymbol{P_{D^2NN}}$. Following the calculation of $\boldsymbol{P_{D^2NN}}$ predicted by the forward model of a trained diffractive permutation network, it was scaled with a multiplicative constant, σ_P, to account for the optical losses:

$$\sigma_P = \frac{\frac{1}{N_i N_o}\sum_{n_i}^{N_i}\sum_{n_o}^{N_o} P_{D^2NN}[n_i, n_o]P[n_i, n_o]}{\frac{1}{N_i N_o}\sum_{n_i}^{N_i}\sum_{n_o}^{N_o} P_{D^2NN}[n_i, n_o]^2}. \tag{6.9}$$

The all-optical transformation error, $\left\|\boldsymbol{P} - \boldsymbol{P_{D^2NN}}\right\|^2$ can be computed based on,

$$\left\|\boldsymbol{P} - \boldsymbol{P_{D^2NN}}\right\|^2 = \frac{1}{N_i N_o}\sum_{n_i}^{N_i}\sum_{n_o}^{N_o}|\sigma_P P_{D^2NN}[n_i, n_o] - P[n_i, n_o]|^2. \tag{6.10}$$

Denoting the lexicographically ordered vectorized version of a 2D input intensity pattern with $I_{in}[s]$, the ground truth output intensity can be found by $\boldsymbol{P}I_{in}[s]$. The PSNR between this ground-truth vector and the output vector synthesized by the forward optical operation of a given, trained diffractive network, $I_{out}[s]$, can be calculated as,

$$PSNR = 20log_{10}\left(\frac{1}{\sqrt{\sum_s |\boldsymbol{P}I_{in}[s] - \sigma I_{out}[s]|^2}}\right), \tag{6.11}$$

where σ is the multiplicative constant defined in Eq. 6.6. The SSIM values were calculated based on the built-in function in TensorFlow, i.e., tf.image.ssim, where the two inputs were 2D versions of $\boldsymbol{P}I_{in}[s]$ and $I_{out}[s]$, representing the ground-truth image and the permuted, all-optical output signal, respectively. All the parameters of tf.image.ssim were taken equal to default values, except that the size of the Gaussian filter was set to be 5×5, instead of 11×11, and the width of the Gaussian filter was set to be 0.75.

Vaccination framework

v-D^{2}NN framework aims to design diffractive optical networks that are resilient against physical error sources, e.g., misalignments, by modeling these factors as random variables and incorporating them into the forward training model. In the training forward model of the vaccinated diffractive networks shown in Fig. 6.4, 4 physical error components were modeled representing the misalignment of each diffractive layer with respect to their ideal location and orientation/angle. The first 3 components represent the statistical variations in the location of each diffractive layer in 3D space. Let the ideal location of a diffractive layer, l, be denoted by the vector $\boldsymbol{X^l} = (x_l, y_l, z_l)$, then at each training iteration i, v-D^{2}NN framework perturbs $\boldsymbol{X^l}$ with a random displacement vector, $\boldsymbol{D^{l,i}} = \left({D_x}^{l,i}, {D_y}^{l,i}, {D_z}^{l,i}\right)$. The components of this 3D displacement vector were defined as uniformly distributed, independent random variables, i.e.,

$$D_x{}^{l,i} \sim U(-\Delta_x, \Delta_x)$$

$$D_y{}^{l,i} \sim U(-\Delta_y, \Delta_y)$$

$$D_z{}^{l,i} \sim U(-\Delta_y, \Delta_y) \tag{6.12}$$

During the training, for each batch of input images, the 3D displacement vector $\boldsymbol{D^{l,i}}$ is updated and accordingly, the location of the layer l is set to be $\boldsymbol{X^{l,i} = X^l + D^{l,i}}$, building up robustness to physical misalignments.

Beyond the displacement of diffractive layers, the physical forward model of a diffractive network is also susceptible to variations in the orientation of the diffractive layers. Ideally, one should include all 3 rotational components, yaw, pitch and roll, however, in this study we only considered the yaw component since in our experimental systems, the pitch and the roll can be controlled with a high precision. The random angle representing the rotation of a diffractive layer l around the optical axis was defined as $D_\theta{}^{l,i} \sim U(-\Delta_{\boldsymbol{\theta}}, \Delta_{\boldsymbol{\theta}})$. With 3 shift components depicted in Eq. 6.12 and the statistical yaw variation modeled through $D_\theta{}^{l,i}$, the vaccinated diffractive networks shown in Fig. 6.4 were trained to build resilience against these 4 misalignment components. The values of $\Delta_{\boldsymbol{x}}$, $\Delta_{\boldsymbol{y}}$, $\Delta_{\boldsymbol{z}}$ and $\Delta_{\boldsymbol{\theta}}$ determining the misalignment tolerance range were defined as a function a common variable v, i.e., $\Delta_{\boldsymbol{x}} = \Delta_{\boldsymbol{y}} = 0.67\lambda v$, $\Delta_{\boldsymbol{z}} = 24\lambda v$ and $\Delta_{\boldsymbol{\theta}} = 4°$.

For the design of the experimentally validated diffractive permutation network, on top of these 4 optomechanical error components (with $v = 0.5$), we also modeled fabrication errors in the form of statistical variations of the material thickness over each diffractive neuron (h). Hence, at a

given iteration, i, the material thickness values over each diffractive unit $h(h_a)$, defined in Eq. 6.8 was perturbed through $h^i(h_a) = h(h_a) + {D_h}^i$, where ${D_h}^i \sim U(-0.025h_m, 0.025h_m)$. Stated differently, the fabricated diffractive layers shown in Fig. 6.5 were designed to be resilient against physical errors on the material thickness values over the diffractive neurons within a range $[-0.0415\lambda, 0.0415\lambda]$.

Chapter 7 All-optical Phase Recovery: Diffractive Computing For Quantitative Phase Imaging

Parts of this chapter have previously been published in D. Mengu et al. “All-optical Phase Recovery: Diffractive Computing For Quantitative Phase Imaging”, Advanced Optical Materials, DOI: 10.1002/adom.202200281. This chapter presents a numerical study that investigates the capabilities of diffractive optical networks in transforming the phase channel of an input objects to intensity for revealing quantitative image.

Quantitative phase imaging (QPI) is a label-free computational imaging technique that provides optical path length information of specimens. In modern implementations, the quantitative phase image of an object is reconstructed digitally through numerical methods running in a computer, often using iterative algorithms. Here, we demonstrate a diffractive QPI network that can perform all-optical phase recovery and synthesize the quantitative phase image of an object by converting the input phase information of a scene into intensity variations at the output plane. A diffractive QPI network is a specialized all-optical processor designed to perform a quantitative phase-to-intensity transformation through passive diffractive surfaces that are spatially engineered using deep learning and image data. Forming a compact, all-optical network that axially extends only ~200-300λ, where λ is the illumination wavelength, this framework can replace traditional QPI systems and related digital computational burden with a set of passive transmissive layers. All-optical diffractive QPI networks can potentially enable power-efficient,

high frame-rate and compact phase imaging systems that might be useful for various applications, including, e.g., on-chip microscopy and sensing.

7.1 Introduction

Optical imaging of weakly scattering phase objects has been of significant interest for decades, resulting in numerous applications in different fields. For example, the optical examination of cells and tissue samples is frequently used in biological research and medical applications, including disease diagnosis. However, in terms of their optical properties, isolated cells and thin tissue sections (before staining) can be classified as weakly scattering, transparent objects[237]. Hence, when they interact with the incident light in an optical imaging system, the amount of light scattered due to the spatial inhomogeneity of the refractive index is much smaller than the light directly passing through, resulting in a poor image contrast at the output intensity pattern. One way to circumvent this limitation is to convert such phase objects into amplitude-modulated samples using chemical stains or tags[238]. In fact, for over a century, histopathology practice has relied on the staining of biological samples for medical diagnosis to bring contrast to various features of the specimen. While these methods generally provide high-contrast imaging (sometimes with molecular specificity), they are tedious and costly to perform, often involving toxic chemicals and lengthy manual staining procedures. Moreover, the use of exogenous stains might cause changes in the physiology of living cells and tissue, creating practical limitations in various biological applications[239].

The phase contrast imaging principle, invented by Frits Zernike, represents a breakthrough (leading to the 1953 Nobel Prize in Physics) on imaging the intrinsic optical phase delay induced by transparent, phase objects without using exogenous agents[240]. Nomarski's differential interference contrast (DIC) microscopy is another method frequently used to investigate phase objects without staining[241]. While both phase contrast imaging and DIC microscopy can offer sensitivity to nanoscale optical path length variations, they reveal the phase information of the

specimen in a *qualitative* manner. On the other hand, quantification and mapping of a sample's phase shift information with high sensitivity and resolution allows for various biomedical applications[242–244]. To address this broad need, quantitative phase imaging (QPI) has emerged as a powerful, label-free approach for optical examination of, e.g., morphology and spatiotemporal dynamics of transparent specimens[239]. The last decades have witnessed the development of numerous digital QPI methods, e.g., Fourier Phase Microscopy (FPM)[245], Hilbert Phase Microscopy (HPM)[246], Digital Holographic Microscopy (DHM)[247–252], Quadriwave Lateral Shearing Interferometer (QLSI)[253] and many others[254–263]. This transformative progress in QPI methods has fostered various applications in, e.g., pathology[248], cell migration dynamics[242,264] and growth[265], immunology[266] and cancer prognosis[267–270], among others[271–278].

A QPI system, in general, consists of an optical imaging instrument based on conventional components such as lenses, beamsplitters, as well as a computer to run the image reconstruction algorithm that recovers the object phase function from the recorded interferometric measurements. In recent years, QPI methods have also benefited from the ongoing advances in machine learning and GPU-based computing to improve their digital reconstruction speed and spatiotemporal throughput[279–284]. For example, it has been shown that feedforward deep neural networks can be used for solving challenging inverse problems in QPI systems, including, e.g., phase retrieval[50,285,286], pixel super-resolution[287] and extension of the depth-of-field[52].

In this work, we report the numerical design of diffractive optical networks[77] to replace digital image reconstruction algorithms used in QPI systems with a series of passive optical modulation surfaces that are spatially engineered using deep learning. The presented QPI diffractive networks (Fig. 7.1) have a compact footprint that axially spans ~240λ and are designed using

deep learning to encode the optical path length induced by a given input phase object into an output intensity distribution that all-optically reveals the corresponding QPI information of the sample. Through numerical simulations, we show that these QPI diffractive network designs can generalize not only to unseen, new phase images that statistically resemble the training image dataset, but also generalize to entirely new datasets with different object features.

It is important to emphasize that these QPI diffractive networks do *not* perform phase recovery from an intensity measurement or a hologram. In fact, the input information is the phase object itself, and the QPI network is trained to convert this phase information of the input scene into an intensity distribution at the output plane; this way, the normalized output intensity image directly reveals the quantitative phase image of the sample in radians.

The diffractive QPI designs reported in this work represent proof-of-concept demonstrations of a new phase imaging concept, and we believe that such diffractive computational phase imagers can find various applications in on-chip microscopy and sensing due to their compact footprint, all-optical computation speed and low-power operation.

7.2 Results

Revealing the optical phase delay induced by an input object by converting or encoding the sample information into an optical intensity pattern at the output plane is a relatively old and well-known technique[240]. Unlike analog phase contrast imaging methods that allow qualitative investigation of the samples, modern QPI systems numerically retrieve the spatial map of the optical phase delay induced by the sample. However, the fundamental idea of encoding the phase information of the object function into the output intensity pattern prevails. For instance, coherent QPI methods use optical hardware, commonly based on conventional optical

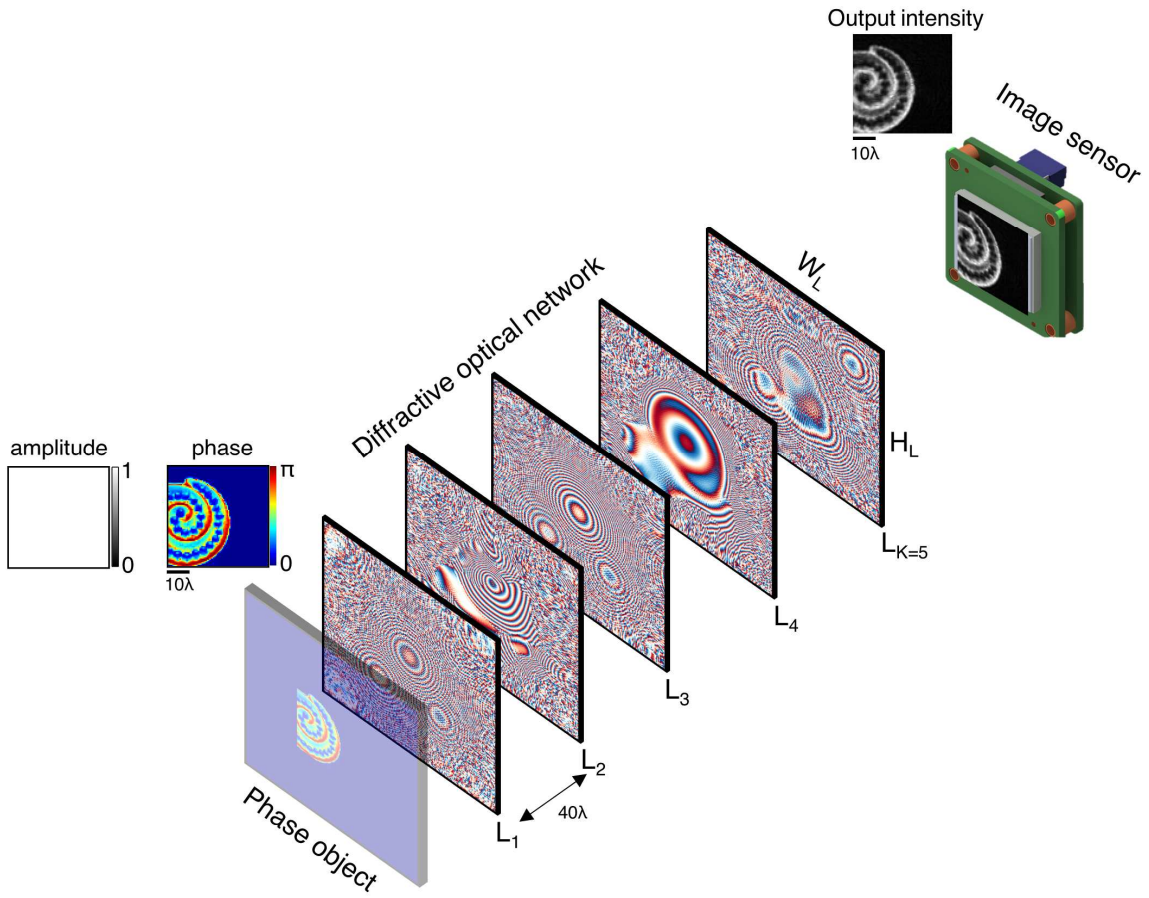

Fig. 7.1 Schematic of a diffractive QPI network that converts the optical phase information of an input object into a normalized intensity image, revealing the QPI information in radians without the use of a computer or a digital image reconstruction algorithm. Optical layout of the presented 5-layer diffractive QPI network, where the total distance between the input and output fields-of-view is 240λ.

components such as lenses and beamsplitters, to generate interference between a reference wave and the object wave over an image sensor-array, creating fringe patterns that implicitly describe the phase function of the input sample. These QPI systems also rely on a phase recovery step implemented in a computer that decodes the object phase information by digitally processing the recorded optical intensity pattern(s), often using iterative algorithms.

To create an all-optical QPI solution without any digital phase reconstruction algorithm, we designed diffractive networks[77,78,166,288,289] that transform the phase information of the input sample into an output intensity pattern, quantitatively revealing the object phase distribution through an intensity recording. Figure 7.1 illustrates the schematic of a 5-layer diffractive network that was trained to all-optically synthesize the QPI signal of a given input phase object (see Methods section for training details). This system can precisely quantify and map the optical path length variations at the input, and unlike the modern QPI systems, it does not rely on a computationally intensive phase reconstruction algorithm or a digital computer.

For a proof-of-concept demonstration, here we considered the design of diffractive QPI networks with unit magnification, such that the input object features in the phase space have the same scale as the output intensity features behind the diffractive network. Since the value of the output optical intensity will depend on external physical factors such as, e.g., the power of the illumination source and the quantum efficiency of the image sensor-array, we used a background region (see Methods section) that surrounds the unit magnification output image to obtain a reference mean intensity. This mean signal intensity value at this background region is used to normalize the output intensity of the diffractive network's image to reveal the quantitative phase information of the sample in radians, i.e., $I_{QPI}(x, y)$ [rad]. Therefore, at the output plane of the

diffractive QPI network, we defined an output signal area that is slightly larger than the input sample field-of-view, where the edges are used to reveal the intensity normalization factor, which makes our diffractive QPI designs ***invariant*** to changes in the illumination beam intensity or the diffraction efficiency of the imaging system, correctly revealing $I_{QPI}(x, y)$, matching the quantitative phase information of the input object in radians.

Figure 7.2a shows the phase-only diffractive layers constituting a diffractive QPI network that is trained using deep learning. In our proof-of-concept numerical experiments, we opted to train and test our diffractive network designs on well-known image datasets to better benchmark the resulting QPI capabilities. Given a normalized greyscale image from a target dataset, $\phi(x, y)$, the corresponding function of a phase object at the input plane can be written as $e^{j\alpha\pi\phi(x,y)}$ where $|\phi(x, y)| \leq 1$. The parameter α determines the range of the phase shift induced by the input object. The diffractive optical network shown in Fig. 7.2a was trained based on $\phi(x, y)$ taken from the Tiny-Imagenet dataset[290] and the parameter, α, was set to be 1 for both training and testing, i.e., $\alpha_{tr} = \alpha_{test} = 1$. Figure 7.2b illustrates the QPI signals, $I_{QPI}(x, y)$, for exemplary test samples from the Tiny-imagenet dataset, never seen by the diffractive network in the training phase, along with the corresponding ground truth images, $\phi(x, y)$. We quantified the success of the QPI signal synthesis performed by the presented diffractive network using the Structural Similarity Index Measure (SSIM)[291] and the peak signal-to-noise ratio (PSNR). The diffractive network shown in Fig. 7.2a provides an SSIM of 0.824±0.050 (mean ± std) and a PSNR of 26.43dB±2.69 over the entire 10K test samples of the Tiny-Imagenet.

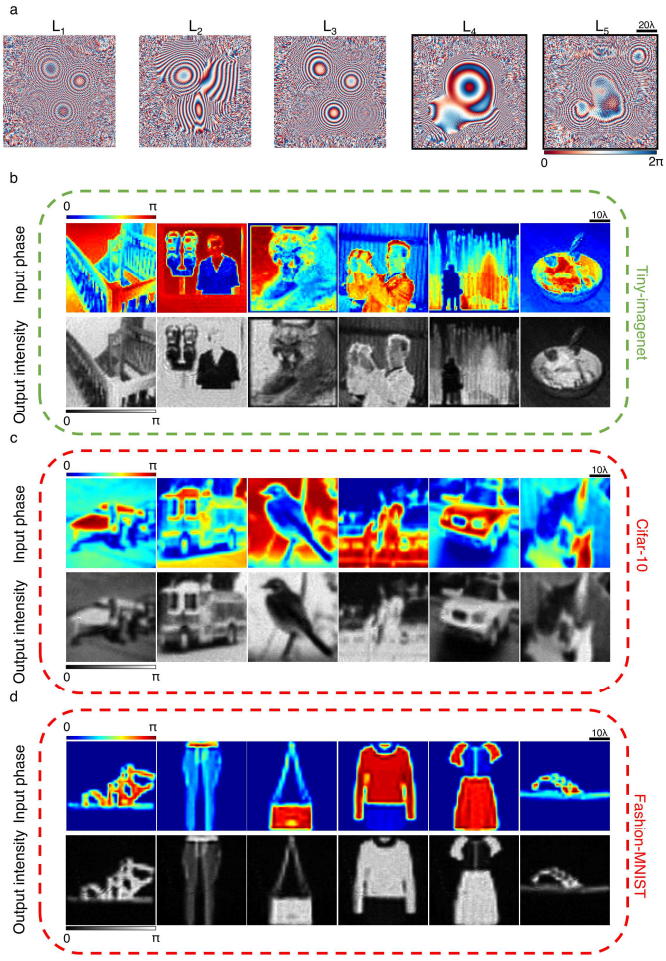

Fig. 7.2 Generalization capability of diffractive QPI networks. a, The phase profiles of the diffractive layers forming the diffractive QPI network trained using phase-encoded images from Tiny-Imagenet dataset, $\phi(x,y)$. b, Exemplary input object images and the corresponding output QPI signals for the test images, never seen by the network during training, taken from the Tiny-Imagenet. Dashed green box indicates that the test images, although not seen by the diffractive network before, belong to the same dataset used in the training. c-d, Same as b, except that the test images are taken from CIFAR-10 and Fashion-MNIST. Dashed red boxes indicate that these test images are from entirely new datasets compared to the Tiny-Imagenet used in the training. The SSIM (PSNR) values achieved by the presented diffractive network are 0.824±0.050 (26.43dB±2.69), 0.917±0.041 (31.98dB±3.15) and 0.596±0.116 (26.94dB±1.5) for the test images from Tiny-Imagenet, CIFAR-10 and Fashion-MNIST datasets, respectively.

Although our diffractive QPI network design can successfully transform the phase information of the samples into quantitative optical intensity information, providing a competitive QPI performance without the need for any digital phase recovery algorithm, one might argue that the underlying phase-to-intensity transformation performed by the diffractive network is data-specific. To shed more light on this, we investigated the generalization capabilities of our diffractive network design by further testing its QPI performance over phase-encoded samples from two completely different image datasets, i.e., CIFAR-10 and Fashion-MNIST, that were not used in the training phase. As shown in Figs. 7.2c-d, the SSIM and PSNR values achieved by the presented diffractive QPI network for quantitative phase imaging of CIFAR-10 (and Fashion-MNIST) images are 0.917±0.041 (and 0.596±0.116) and 31.98dB±3.15 (and 26.94dB±1.5), respectively. Interestingly, the QPI signal synthesis quality turned out to be higher for CIFAR-10 images compared to the performance of the same diffractive network on the Tiny-Imagenet test samples, even though CIFAR-10 has an entirely different set of objects and spatial features (which were never used during the training phase). This could be partially attributed to the difference in the original size of the Tiny-Imagenet (64×64-pixel) and CIFAR-10 (32×32-pixel) images. Considering that the physical dimensions of the input field-of-view in our network configuration is 42.4λ×42.4λ, the size of the smallest spatial feature becomes $\frac{42.4\lambda}{64} = 0.6625\lambda$ and $2 \times 0.6625\lambda$ for Tiny-Imagenet and CIFAR-10 datasets, respectively; this makes CIFAR-10 test samples relatively easier to image through the diffractive QPI network.

Next, we numerically quantified the smallest resolvable linewidth and the related phase sensitivity of our diffractive QPI network design using binary phase gratings as test objects (see Fig. 7.3). Such resolution test targets were not used as part of the training, which only included the Tiny-Imagenet dataset. The presented diffractive network performs QPI with diffractive

layers of size 106λ×106λ that are placed 40λ apart from each other and the input/output fields-of-view (see Fig. 7.1). This physical configuration reveals that the numerical aperture (NA) of our diffractive network is $\sin\left(tan^{-1}\left(\frac{106\lambda}{2\times 40\lambda}\right)\right) = \sim 0.8$, which corresponds to a diffraction-limited resolvable linewidth of 0.625λ. Our numerical analysis in Fig. 7.3a showed that the smallest resolvable linewidth with our diffractive QPI design was ~0.67λ, when the input gratings were 0-π encoded, closely matching the resolvable feature size determined by the NA of our system; also note that the effective feature size of the training samples from Tiny-Imagenet is 0.6625λ. This analysis means that our training phase was successful in approximating a general-purpose quantitative phase imager despite using relatively lower resolution training images, coming close to the theoretical diffraction limit imposed by the physical structure of the diffractive QPI network.

The input phase contrast is another crucial factor affecting the resolution of QPI achieved by our diffractive network design. To shed more light on this, we numerically tested our diffractive QPI network on binary gratings with two different linewidths, 0.67λ and 0.75λ, at varying levels of input phase contrast, as shown in Fig. 7.3b. Based on the resulting diffractive QPI signals illustrated in Figs. 7.3a-b, the 0.67λ linewidth grating remains resolvable until the input phase contrast falls below 0.25π. The last column of Fig. 7.3b suggests that when the contrast parameter (α_{test}) is taken to be 0.1, the noise level in the QPI signal generated by the diffractive network increases to a level where the 0.67λ linewidth grating cannot be resolved anymore. On the other hand, 0.75λ linewidth grating remains to be partially resolvable despite the noisy background, even at 0-0.1π phase contrast (i.e., $\alpha_{test} = 0.1$).

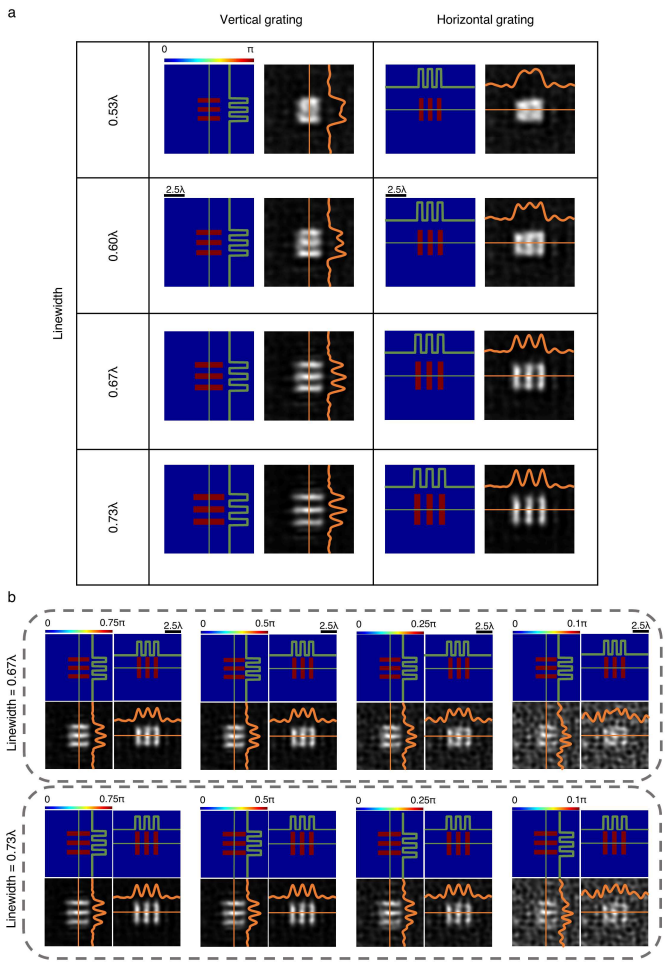

Fig. 7.3 Spatial resolution and phase sensitivity analysis. Input phase image and the corresponding output diffractive QPI signal for binary, 0-π phase encoded grating objects. The diffractive QPI network can resolve features as small as ~0.67λ. b, Analysis of the relationship between the input phase contrast and the resolvable feature size. The diffractive QPI network can resolve 0.67λ linewidth for a phase encoding range that is larger than 0.25π. Below this phase contrast, the resolution slowly degrades; for example, at 0-0.1π phase encoding, the background noise shadows the QPI signal of the grating with a linewidth of 0.67λ, while a larger linewidth (0.73λ) grating is still partially resolvable.

We also conducted a similar analysis on the effect of the input phase contrast over the quality of QPI performed by the presented diffractive network. By setting the phase contrast parameter α_{test} to 9 different values between 0.1 and 2.0 for all three image datasets (Tiny-Imagenet, CIFAR-10 and Fashion-MNIST), we quantified the resulting SSIM and PSNR values for the reconstructed images at the output plane of the diffractive QPI network. Figures 7.4a-c illustrate the mean and standard deviations of the SSIM and PSNR metrics as a function of α_{test} for all three image datasets. A close examination of Figs. 7.4a-c reveals that both SSIM and PSNR peaks at $\alpha_{test} = 1$, which matches the phase encoding range used during the training phase, i.e., $\alpha_{tr} = \alpha_{test} = 1$. To the left of these peaks, where $\alpha_{test} < \alpha_{tr} = 1$, there is a slight degradation in the performance of the presented diffractive QPI network, mainly due to the increasing demand in phase sensitivity at the resulting image, $I_{QPI}(x, y)$. With $\alpha_{tr} = 1$ and 8-bit quantization of input signals, the phase step size that the diffractive QPI network was trained with was $\frac{\pi}{256} = 0.0123$ radians; however, when α_{test} deviates from the training, for instance $\alpha_{test} = 0.5$, then the smallest phase step size that the diffractive network is tasked to sense becomes $\frac{0.5\pi}{256} = 0.0062$ radians. In other words, the diffractive network must be 2× more phase sensitive compared to the level it was trained for, causing some degradation in the SSIM and PSNR values as shown in Figs. 7.4a-c for $\alpha_{test} < \alpha_{tr} = 1$.

On the other hand, when the input phase encoding exceeds the [0, π] range used during the training phase, the degradation in diffractive QPI signal quality is more severe. As α_{test} approaches to 2.0, the errors and artifacts created by the presented diffractive network in computing the QPI signal increase. Interestingly, at $\alpha_{test} = 1.99$, the forward optical transformation of the diffractive QPI network starts to act as an edge detector. A straightforward

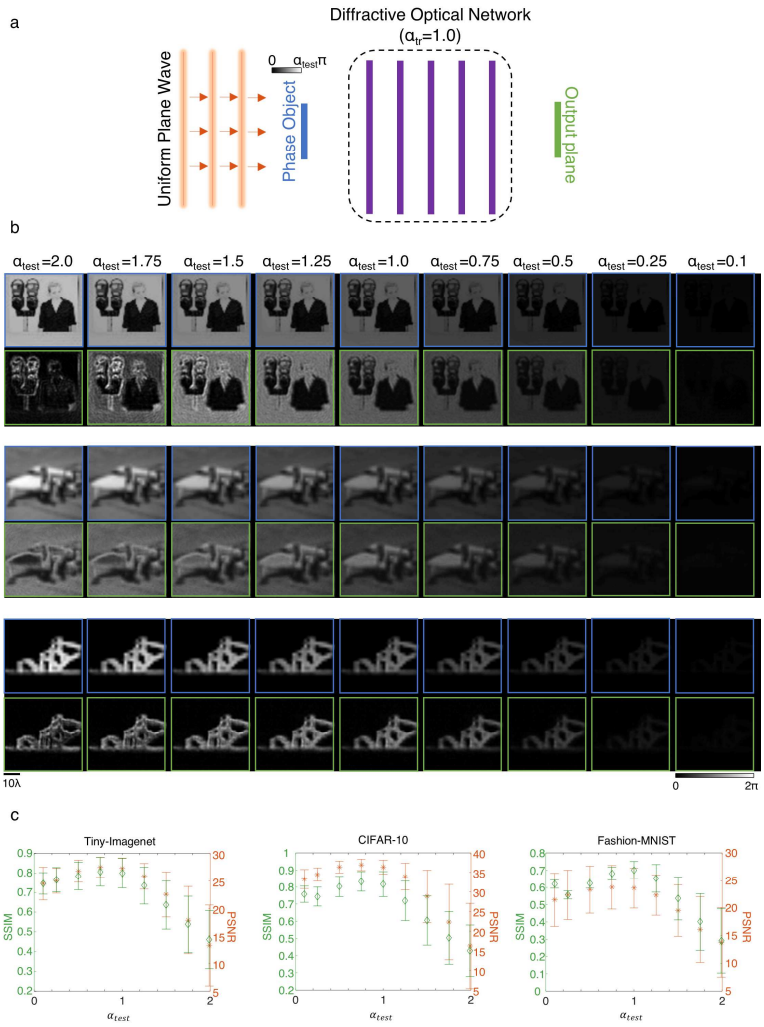

Fig. 7.4 The impact of input phase range on the diffractive QPI signal quality. a, A schematic of the diffractive QPI network that was trained with α_tr=1.0, meaning that the training images had [0 : π] phase range. b, Pairs of ground-truth input phase images (top rows) and the diffractive QPI signal (bottom rows) for different images taken from Tiny-Imagenet (top), CIFAR-10 (middle) and Fashion-MNIST (bottom), at different levels of phase encoding ranges dictated by (from left-to-right) α_test=2, α_test=1.75, α_test=1.5, α_test=1.25, α_test=α_tr=1.0, α_test=0.75, α_test=0.5, α_test=0.25, α_test=0.1. c, The SSIM and PSNR values of the diffractive QPI signals with respect to the ground-truth images as a function of α_test.

solution to mitigate this performance degradation is to train the diffractive network with $\alpha_{tr} = 2.0 - \epsilon$, where ϵ is a small number, meaning that during the training phase, the dynamic range of the phase values at the input plane will be within [0, 2π). Figure 7.5 illustrates an example of this for a 5-layer diffractive QPI network that was trained with $\alpha_{tr} = 1.99$. This new diffractive network has the same physical layout and architecture as the previous one shown in Fig. 7.2. The only difference between the two diffractive QPI networks is the phase range covered by the input samples used during their training ($\alpha_{tr} = 1.0$ vs. $\alpha_{tr} = 1.99$). Since the design evolution of this new diffractive QPI network is driven by input samples covering the entire [0, 2π) phase range, in the case of $\alpha_{test} = \alpha_{tr} = 1.99$, it provides a much better QPI performance compared to the diffractive network shown in Fig. 7.2. This improved diffractive QPI performance can also be visually observed by comparing the images shown in Fig. 7.4 and Fig. 7.5 under the $\alpha_{test} = 1.99$ column.

7.3 Discussion

Compared to earlier works on diffractive optical networks that demonstrated amplitude imaging[77], the presented QPI diffractive networks report significant advances. While a conventional amplitude imaging task requires the diffractive network to achieve a point-to-point intensity mapping between the input and output fields-of-view, all-optical synthesis of the QPI signal describing the phase variations of an input object is a nonlinear operation as it converts the input phase information into quantitative output intensity variations, and this nonlinear operation (phase-to-intensity transformation) is all-optically approximated through our QPI diffractive networks, the magnitude-squared signal detection on the opto-electronic sensor and the subsequent normalization step depicted in Eq. 7.4. Furthermore, a vital feature of the presented diffractive QPI networks is that their operation is invariant to changes in the input beam intensity

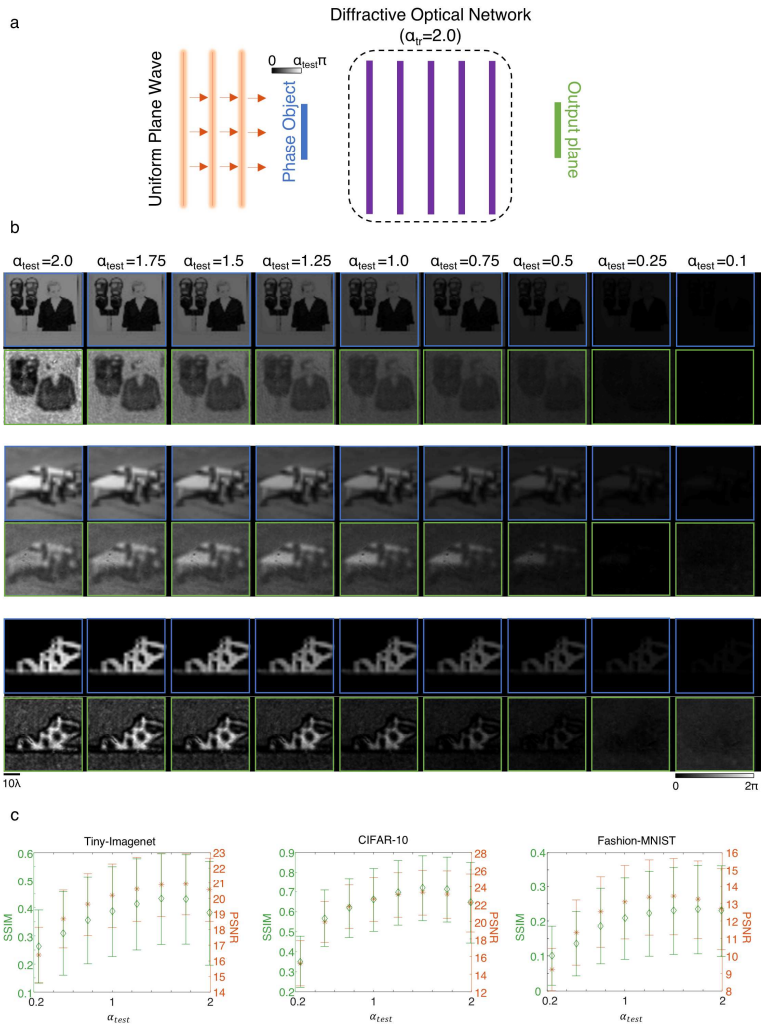

Fig. 7.5 The impact of input phase range on the diffractive QPI signal quality. Same as Fig. 7.4, except that this diffractive QPI network was trained with α_tr=2.0, meaning that the training images had [0, 2π) phase range, instead of [0, π].

or the power efficiency of the diffractive detection system; by using the mean intensity value surrounding the output image field-of-view as a normalization factor, the resulting diffractive image intensity $I_{QPI}(x, y)$ reports the phase distribution of the input object in radians. Moreover, the presented diffractive optical networks are composed of passive layers, and therefore perform QPI without any external power source other than the illumination light. It is true that the training stage of a diffractive QPI network takes a significant amount of time (e.g., ~40 hours) and consumes some energy for training-related computing. But this is a one-time training effort, and in the image inference stage, there is no power consumption per object (except for the illumination), and the reconstructed image reveals the quantitative phase information of the object at the speed of light propagation through a passive network, without the need for a graphics processing unit (GPU) or a computer. One should think of a diffractive network's design, training and fabrication phase (a one-time effort) similar to the design/fabrication/assembly phase of a digital processor or a GPU that we use in our computers.

Another important aspect of the presented diffractive QPI framework is its generalization capability over image datasets other than the one used in the training phase, as shown in Fig. 7.2. To further test the role of the training dataset in the generalization capability of the diffractive QPI system, we trained a new diffractive network with a physical architecture identical to that of the QPI diffractive network shown in Fig. 7.2. The only difference was that this new diffractive optical network was trained using the Fashion-MNIST dataset instead of the Tiny-Imagenet. Compared to the QPI diffractive network shown in Fig. 7.2 (trained with Tiny-Imagenet) that achieved (SSIM, PSNR) performance metrics of (0.824±0.050, 26.43dB±2.69), (0.917±0.041, 31.98dB±3.15) and (0.596±0.116, 26.94dB±1.5) for Tiny-Imagenet, CIFAR-10 and Fashion-MNIST test datasets, respectively, this new QPI diffractive network (trained with Fashion-

MNIST) provided (SSIM, PSNR) performance metrics of (0.622±0.085, 19.97dB±2.36), (0.699±0.106, 21.38dB±2.7) and (0.816±0.060, 31.26dB±2.12), for the same test datasets, respectively. From this comparison, we can conclude that: (1) the QPI diffractive network can be trained with other image datasets and successfully generalize to achieve phase recovery for new types of test images, and (2) the richness of the phase variations in the training images impacts the performance and generalization capability of the QPI diffractive network; for example, the QPI diffractive network trained with Tiny-Imagenet achieved relatively better generalization to new phase images obtained from CIFAR-10 test dataset when compared to the QPI diffractive network trained with Fashion-MNIST. To further quantify the generalization performance of the presented QPI diffractive network shown in Fig. 7.2 (trained with Tiny-Imagenet), we blindly tested it with phase images of thin Pap (Papanicolaou) smear samples as shown in Fig. 7.6. Although, this QPI diffractive optical network was only trained using the phase-encoded images from Tiny-Imagenet, it very well generalized to new types of samples, performing quantitative phase retrieval and QPI on the phase images of Pap smear samples, with output SSIM and PSNR values of 0.663±0.047 and 25.55dB±1.44, respectively (see Fig. 7.6).

The output power efficiency of the presented QPI networks is mainly affected by two factors: diffraction efficiency of the resulting network and material absorption. In this study, we assumed the optical material of diffractive surfaces has a negligible loss for the wavelength of operation, similar to the properties of optical glasses, e.g., BK-7, in the visible part of the spectrum. Beyond the material absorption, another possible source of power loss in a physically implemented diffractive network is the surface back-reflections, which might potentially be minimized through e.g., anti-reflection thin-film coatings[292]. For example, the diffractive QPI network reported in Fig. 7.2 achieves ~2.9% mean diffraction efficiency for the entire 10K test set of

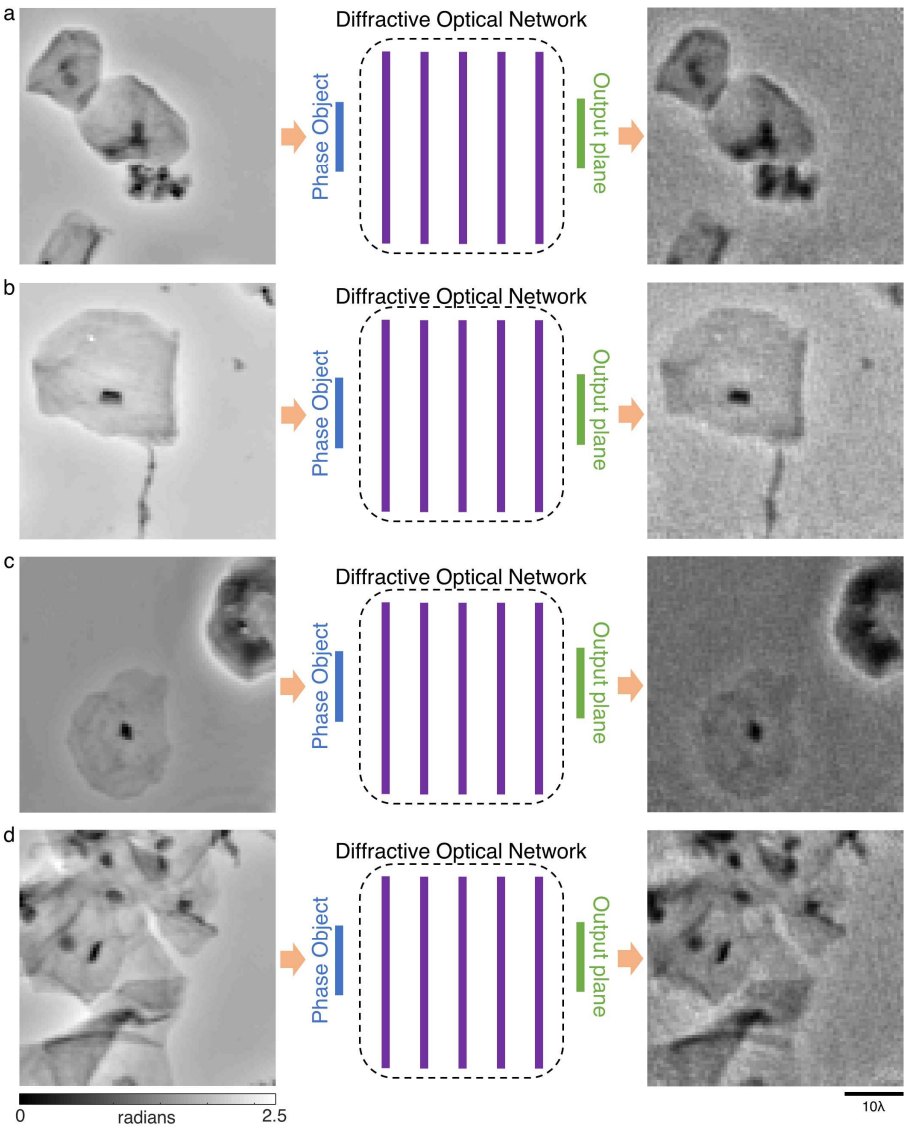

Fig. 7.6 The signal synthesis performance of a QPI diffractive optical network on Pap-smear samples. The input images represent the phase channel of Pap-smear samples (monolayer of cells), and the QPI signals (output intensity) are synthesized by the diffractive optical network shown in Fig. 7.2. Although, this QPI diffractive network model is trained using only the images from Tiny-imagenet, it can blindly achieve SSIM and PSNR values of 0.663±0.047 and 25.55dB±1.44, respectively, over these Pap-smear samples.

Tiny-Imagenet. It is important to note that during the training of this diffractive QPI network, the training cost/loss function was purely based on decreasing the QPI errors at the output plane, and there was no other loss term or regularizer to enforce a more power-efficient operation. In fact, by including an additional loss term for regulating the balance between the QPI performance and diffraction efficiency (see Methods section), we demonstrated that it is possible to design more efficient diffractive QPI networks with a minimal compromise on the output image quality; see Fig. 7.7, where all the diffractive network designs share the same physical layout shown in Fig. 7.1. For example, a more efficient diffractive QPI network design with 6.31% power efficiency at the output plane offers QPI signal quality with an SSIM of 0.815±0.0491. Compared to the original diffractive QPI network design that solely focuses on output image quality, the SSIM value of this new diffractive network has a negligible decrease while its diffraction efficiency at the output plane is improved by more than 2-fold. Further shifting the focus of the QPI network training towards improved power efficiency can result in a solution that can synthesize QPI signals with >11% output diffraction efficiency, also achieving an SSIM of 0.771±0.0507 (see Fig. 7.7). We should note here that a standard phase contrast microscope also contains some filters, apertures, lenses and other optical components that block and/or scatter the sample light, all of which also cause some power loss. However, such conventional optical components have very well established fabrication technologies supporting their optimized use in a microscope design. With advances in diffractive optical computing, more efficient diffractive surface designs[293] can be enabled in the future to further increase the output diffraction efficiencies of diffractive networks.

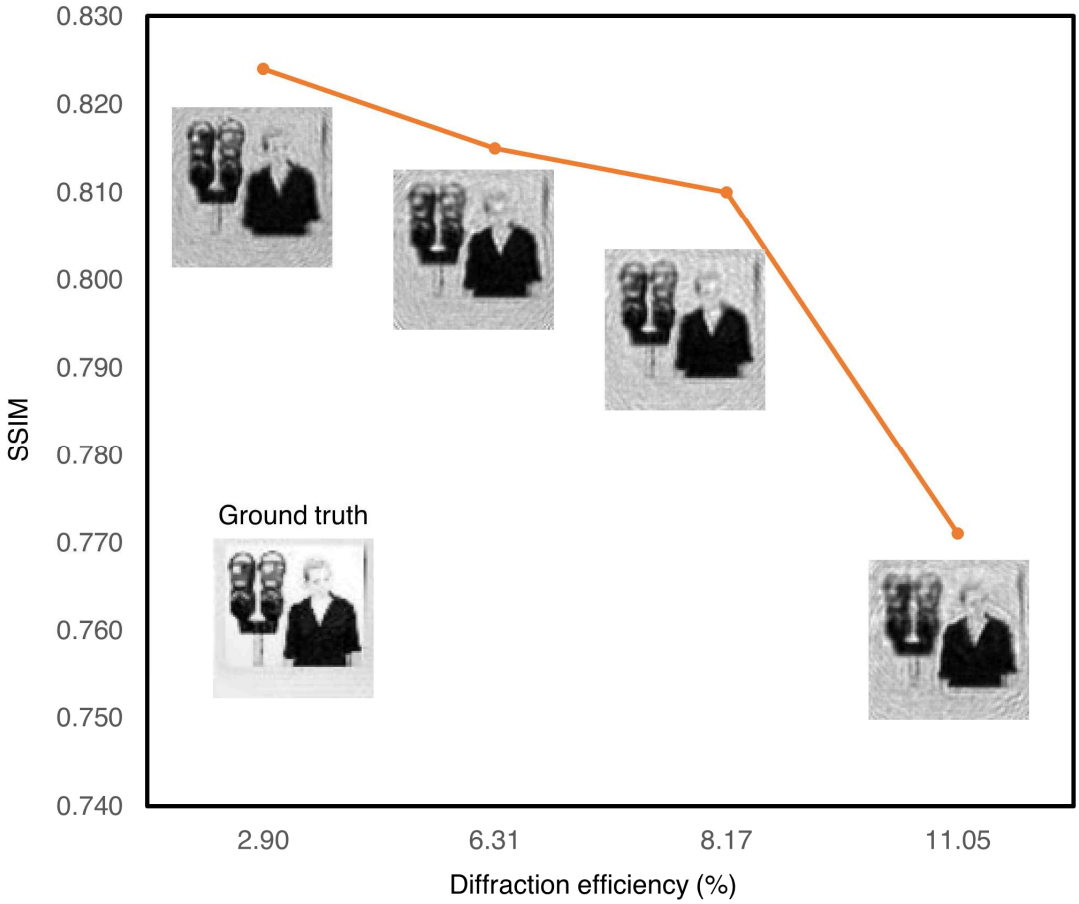

Fig. 7.7 Diffractive QPI signal quality and the power efficiency trade-off. We report 4 different diffractive QPI network models trained using [0 : π] phase-encoded samples from the Tiny-Imagenet dataset. The SSIM on the y-axis reflects the mean value computed over the entire 10K test images of the Tiny-Imagenet dataset. The diffractive QPI network that provides the highest SSIM is the network shown in Fig. 7.2, which was trained solely based on the structural loss function (Eq. 7.5) totally ignoring the diffraction efficiency of the resulting solution. The loss function used for the training of the other 3 diffractive QPI networks includes a linear superposition of the structural loss function (Eq. 7.5) and the diffraction efficiency penalty term depicted in Eq. 7.7. The multiplicative constant γ which determines the weight of the diffraction efficiency penalty was taken as 0.1, 0.4 and 5.0 for these 3 diffractive QPI networks, providing an output diffraction efficiency of 6.31% , 8.17% and 11.05%, respectively.

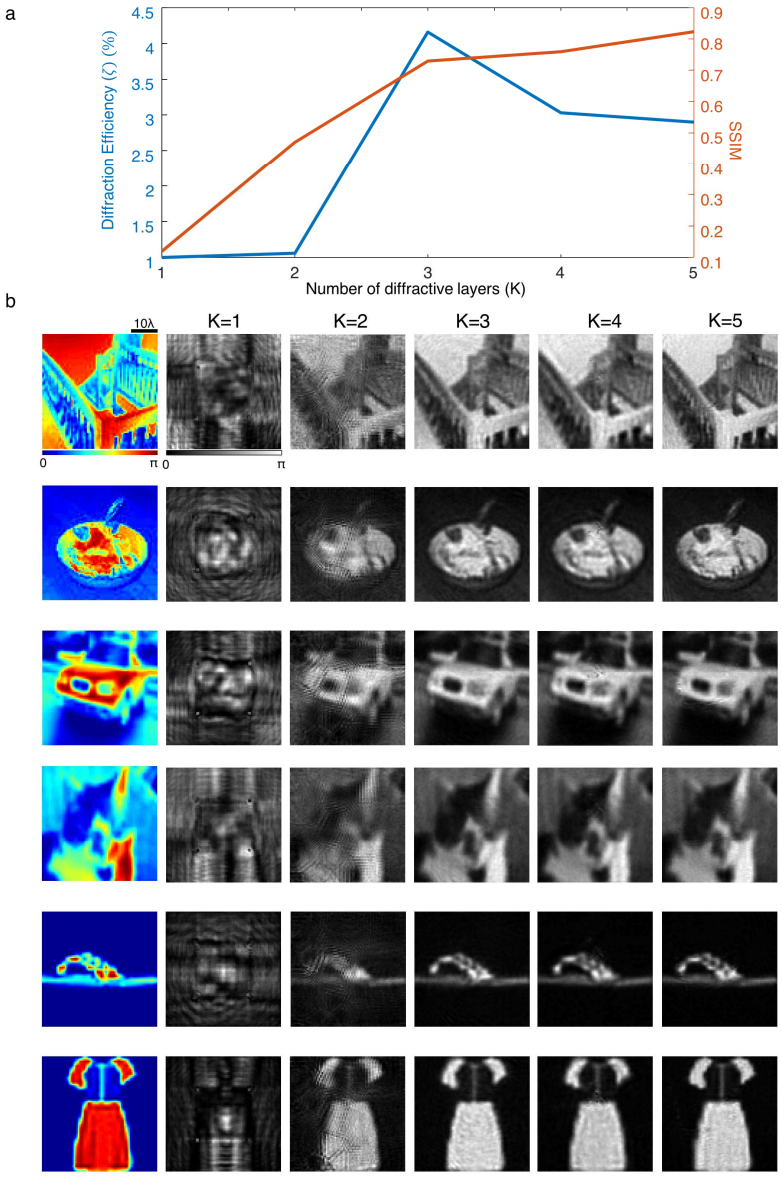

Fig. 7.8 The impact of the number (K) of trainable layers on the diffractive QPI signal quality and the output diffraction efficiency. a, Output diffraction efficiency and SSIM as a function of K. There are 5 different diffractive QPI network designs reported here, with K=1,2,3,4 and 5 trainable, phase-only diffractive surfaces; the K=5 diffractive QPI network is the same one shown in Fig. 7.2. b, Exemplary input object images from Tiny-Imagenet (top-2 rows), CIFAR-10 (middle-2 rows) and Fashion-MNIST (bottom-2 rows) and the corresponding output QPI signals synthesized by the diffractive QPI network designs with K=1,2,3,4 and 5.

Another crucial parameter in a diffractive network design is the number of diffractive layers within the system; Figure 7.8 illustrates the results of our analysis on the relationship between the diffractive QPI performance and the number of diffractive layers within the system. It has previously been shown through both theoretical and empirical evidence that deeper diffractive optical networks can compute an arbitrary complex-valued linear transformation with lower approximation errors, and they demonstrate higher generalization capacity for all-optical statistical inference tasks[172,294]. Figure 7.8 confirms the same behavior: improved QPI performance is achieved by increasing the number of diffractive layers, K. When K=1, the trained diffractive network fails to compute the QPI signal for a given input phase object, as evident from the extremely low SSIM values and the exemplary images shown in Fig. 7.6b. On top of that, the diffraction efficiency is also very low, ~1%, with a single-layer diffractive network configuration (K=1). With K=2 trainable diffractive surfaces, the diffraction efficiency stays very low, while the QPI signal quality improves. When we have K=3 diffractive layers in our QPI network design, we observe a significant improvement in both the diffraction efficiency and the output SSIM compared to K=1 or 2. Beyond K=3, the structural quality of the output QPI signal keeps improving as we add more layers to the diffractive network architecture. However, this improvement does not translate into better diffraction efficiency as the training loss function does not include a power efficiency penalty term. Earlier results reported in Fig. 7.7 clearly show the impact of adding such a regularizer term in the training loss function for improving the diffraction efficiency of the QPI network, reaching >11% power efficiency with a minor sacrifice in the structural fidelity of the output images.

It is also important to note that as the number of diffractive layers increases, the system (if the diffractive network is *not* trained accordingly) becomes more sensitive to physical

misalignments that might be induced through e.g., fabrication and/or opto-mechanical errors[103]. To shed further light on this, we tested the sensitivity of the QPI diffractive network shown in Fig. 7.2 against axial misalignments of the sensor array at the output plane with respect to the diffractive layers. Although, the SSIM and PSNR values of the all-optical QPI signal exhibit a decrease when the output image sensor is placed at a different axial location than the correct position assumed in the design of the QPI diffractive network. However, one can introduce misalignment resilient diffractive designs with the incorporation of "vaccination" in the training of the diffractive network, where such misalignments are randomly introduced during the training process to guide the optimization of the diffractive surfaces to build resilience toward uncontrolled misalignments[103]. For example, using this vaccination strategy, it has been shown that diffractive networks can be trained to provide an extended depth-of-field, mitigating performance degradation due to object and/or sensor plane misalignments[295,296]. The incorporation of such vaccination methods into the training stage of diffractive QPI networks would in general result in more robust designs against misalignments. Beyond misalignments, another practical issue regarding the implementation of diffractive QPI systems that needs to be discussed is the bit depth of the phase modulation on the diffractive layers. During the training of the QPI diffractive networks, it was assumed that the phase modulation over a diffractive surface can take any value in the range $[0,2\pi)$.

Although, the diffractive networks analyzed and presented in this study are designed to achieve the QPI task with a unit magnification, this is not a limitation of the underlying framework. Depending on the targeted spatial resolution, imaging field-of-view and throughput, diffractive QPI systems with a magnification larger than 1 can also be devised according to the pixel size and the active area of a desired focal-plane-array at the output plane. With the wide

availability of modern CMOS image sensor technology that has sub-micron pixel sizes, unit magnification imaging systems provide a fine balance between the sample field-of-view and the spatial resolution that can be achieved; therefore, unit magnification imaging systems enable compact and chip-scale microscopy tools that provide a substantial increase in the sample field-of-view and volume that can be probed with a decent spatial resolution[250].

In summary, the presented diffractive QPI networks convert the phase information of an input object into an intensity distribution at the output plane in a way that the normalized output intensity reveals the phase distribution of the object in radians. Being resilient to input light intensity variations and power efficiency changes in the diffractive set-up, this QPI network can replace the bulky lens-based optical instrumentation and the computationally intensive reconstruction algorithms employed in QPI systems, potentially offering high-throughput, low-latency, compact and power-efficient QPI platforms which might fuel new applications in on-chip microscopy and sensing. In addition, depending on the application, they can also be trained to all-optically perform various machine learning tasks (e.g., image segmentation [297] and phase unwrapping) using the phase information channel describing transparent input objects; they can also be integrated with electronic back-end neural networks to enable multi-task, resource-efficient hybrid machine learning systems [78,166]. Fabrication and assembly of such diffractive QPI systems operating in the visible and near IR wavelengths can be achieved using two-photon polymerization-based 3D printing methods as well as optical lithography tools[231,298,299].

7.4 Methods

Optical forward model of diffractive QPI networks

The optical wave propagation in air, between successive diffractive layers, was formulated based on the Rayleigh-Sommerfeld diffraction equation. According to this formulation, the free-space propagation inside a homogeneous and isotropic medium is modeled as a shift-invariant linear system with the impulse response,

$$w(x,y,z) = \frac{z}{r^2}\left(\frac{1}{2\pi r} + \frac{n}{j\lambda}\right)\exp(\frac{j2\pi nr}{\lambda}) \tag{7.1}$$

where $r = \sqrt{x^2 + y^2 + z^2}$. In Eq. 7.1, the parameters n and λ denote the refractive index of the medium ($n = 1$ for air), and the wavelength of the illumination light, respectively. Accordingly, a diffractive neuron, i, located at (x_i, y_i, z_i) on k^{th} layer can be considered as the source of a secondary wave, $u_i^k(x,y,z)$,

$$u_i^k(x,y,z) = w_i(x,y,z)t(x_i,y_i,z_i)\sum_{q=1}^{N} u_q^{k-1}(x_i,y_i,z_i) \tag{7.2}.$$

where the summation in Eq. 7.2 represents the field generated over the diffractive neuron located at (x_i, y_i, z_i) by the neurons on the previous, $(k-1)^{th}$, layer. From Eq. 7.1, the function $w_i(x,y,z)$ in Eq. 7.2 can be written as,

$$w_i(x,y,z) = \frac{z - z_i}{r^2}\left(\frac{1}{2\pi r} + \frac{n}{j\lambda}\right)\exp(\frac{j2\pi nr}{\lambda}) \tag{7.3},$$

with $r = \sqrt{(x-x_i)^2 + (y-y_i)^2 + (z-z_i)^2}$. The multiplicative term $t(x_i, y_i, z_i)$ in Eq. 7.2 denotes the transmittance coefficient of the neuron, i, which, in its general form, can be written as, $t(x_i, y_i, z_i) = a_i \exp(j\theta_i)$. Depending on the diffractive layer fabrication method and the related optical materials, both a_i and θ_i might be a function of other physical parameters, e.g.,

material thickness in 3D printed diffractive layers and driving voltage levels in spatial light modulators. In earlier works on diffractive networks[77,80,103,166], it has been shown that it is possible to directly train such physical parameters through deep learning. On the other hand, a more generic way of optimizing a diffractive network is to define the amplitude a_i and θ_i as learnable parameters. In this study, we constrained our analysis to phase-only diffractive surfaces where the amplitude coefficients, a_i, were all taken as 1 during the entire training. Thus, the only learnable parameters of the presented diffractive networks are the phase shifts applied by the diffractive features, θ_i. For all the diffractive networks that we trained, the initial value of all θ_is was set to be 0, i.e., the initial state of a diffractive network (before the training kicks in) is equal to the free-space propagation of the input light field onto the output plane.

The design of diffractive QPI network

During our deep learning-based diffractive network training, we sampled the 2D space with a period of 0.53λ, which is also equal to the size of each diffractive feature ('neuron') on the diffractive surfaces. Although we described the forward optical model over continuous functions in the previous subsection, training of the presented diffractive networks was performed using digital computers. Hence, we denote the input and output signals using their discrete counterparts for the remaining part of this sub-section with a spatial sampling period of 0.53λ in both directions (*x* and *y*). In the physical layout of the presented diffractive optical networks, the size of the input field-of-view was set to be 42.4λ×42.4λ, which corresponds to 80×80 2D vectors defining the phase distributions of input objects. With $\boldsymbol{I}[m,n]$ denoting an image of size $M \times N$ from a dataset, we applied 2D linear interpolation to compute the 2D vector $\boldsymbol{\phi}[q,p]$ of size 80×80. Note that the values of M and N depend on the used image dataset. Specifically, for

Tiny-Imagenet $M = N = 64$, while for CIFAR-10 and Fashion-MNIST datasets, $M = N = 32$ and $M = N = 28$, respectively. The scattering function within the input field-of-view of the diffractive networks was defined as a pure phase function (see Fig. 7.1) in the form of $e^{j\alpha\pi\boldsymbol{\phi}[q,p]}$.

The physical dimensions of each diffractive layer were set to be 106λ on both *x* and *y* axes, i.e., each diffractive layer contains 200×200 = 40K neurons. For instance, the 5-layer diffractive network shown in Fig. 7.2 has 0.2 million neurons, and hence 0.2 million trainable parameters, θ_i, $i = 1,2,\ldots,0.2 \times 10^6$. In our forward optical model, we set all the distances between (1) the first diffractive layer and the input field-of-view, (2) two successive diffractive layers, and (3) the last diffractive layer and the output plane, as 40λ resulting in an NA of ~0.8. With the size of each diffractive feature/neuron taken as 0.53λ, the diffraction cone angle of the secondary wave emanating from each neuron ensures optical communication between all the neurons on two successive surfaces (axially separated by 40λ), while also enabling a highly compact diffractive QPI network design. For instance, the total axial distance from the input field-of-view to the output plane of a 5-layer diffractive QPI network shown in Fig. 7.1 is only ~240λ.

The size of the QPI signal area at the output plane including the reference/background region was set to be 43.56λ×43.56λ, i.e., the reference region extends on both directions on *x* and *y* axes by 0.53λ, (43.56λ=42.4λ+2×0.53λ). If we denote the background optical intensity over this reference region as $\boldsymbol{I_R}[r]$ and the optical intensity within the QPI signal region as $\boldsymbol{I_S}[q,p]$, then according to our forward model, $\boldsymbol{I_{QPI}}[q,p]$ is found by,

$$\boldsymbol{I_{QPI}}[q,p] = \frac{\boldsymbol{I_S}[q,p]}{B}, \tag{7.4}$$

where $B = \frac{1}{N_R}\sum_{r=1}^{N_R} \boldsymbol{I_R}[r]$ is the mean background intensity value, N_R denotes the number of discretized intensity samples within the reference region. According to Eq. 7.4, for a given input object/sample, the final diffractive QPI signal, $\boldsymbol{I_{QPI}}[q,p]$ reports the output phase image in radians.

To guide the evolution of the diffractive layers according to the QPI signal in Eq. 7.4, at each iteration of the deep learning-based training of the presented diffractive QPI networks, we updated the phase parameters, θ_i, using the following normalized mean-squared-error[300],

$$\mathcal{L} = \frac{1}{N_R + N_S} \sum_{l=1}^{N_R+N_S} |\boldsymbol{o}[l] - \sigma \boldsymbol{o}'[l]|^2, \tag{7.5}$$

where, N_s is the total number of discretized samples representing the QPI signal area, i.e., $N_s = 80 \times 80$. The vectors $\boldsymbol{o}$ and $\boldsymbol{o}'$ are 1D counterparts of the associated 2D discrete signals, $\boldsymbol{o}[q,p]$ and $\boldsymbol{o}'[q,p]$, computed based on lexicographically ordered vectorization operator. They denote the ground-truth phase signal of the input object and the diffractive intensity signal synthesized by the QPI network at a given iteration, respectively. Both the ground truth vector, $\boldsymbol{o}$, and $\boldsymbol{o}'$ cover the output sample field-of-view and the reference signal region surrounding it, hence their size is equal to $N_R + N_S = 82 \times 82$. The 2D vector $\boldsymbol{o}[q,p]$ is defined based on the input vector $\boldsymbol{\phi}[q,p]$. First, we equalize the size of the two vectors by padding the 80×80 vector $\boldsymbol{\phi}[q,p]$ to the size 82×82. The values over the padded region are equal to $\frac{1}{\alpha\pi}$. This padded vector was then scaled with the multiplicative constant $\alpha\pi$ such that the 80×80 part in the middle represents the argument of the phase function $e^{j\alpha\pi\boldsymbol{\phi}[q,p]}$. The reference signal region surrounding this 80×80 part has all ones, implying that the mean intensity over this area will correspond to

1 rad. By computing the loss function in Eq. 7.5 based on a ground-truth vector that also includes the desired reference signal intensity, we implicitly enforce/train the diffractive QPI network to synthesize a uniformly distributed intensity over the reference signal area, although this is not a requirement for the QPI networks' operation.

The multiplicative term, σ, in Eq. 7.5 is a normalization constant that was defined as[300],

$$\sigma = \frac{\frac{1}{N_R + N_S}\sum_{l=1}^{N_R+N_S} \boldsymbol{o}[l]\boldsymbol{o}'^{*}[l]}{\frac{1}{N_R + N_S}\sum_{l=1}^{N_R+N_S} |\boldsymbol{o}'[l]|^{\mathbf{2}}}, \tag{7.6}$$

The structural loss function, $\mathcal{L}$, in Eq. 7.5 drives the QPI quality, and it was the only loss term used during the training of the diffractive networks shown in Figs. 7.2, 7.5 and 7.8. The training of the diffractive network designs with output diffraction efficiencies of ≥ 2.9% shown in Fig. 7.7, on the other hand, use a linear mix of the structural loss in Eq. 7.5 and an additional loss term penalizing poor power efficiency, i.e., $\mathcal{L}' = \mathcal{L} + \gamma\mathcal{L}_p$. The functional form of the power efficiency-related penalty $\mathcal{L}_p$ was defined as,

$$\mathcal{L}_p = e^{-\eta}, \tag{7.7}$$

where η stands for the percentage of power efficiency,

$$\eta = \frac{P_{out}}{P_1} \times 100, \tag{7.8}$$

with P_1 denoting the optical power incident on the 1st diffractive layer and $P_{out} = \sum_{l=1}^{N_R+N_S} |\boldsymbol{o}'[l]|^{\mathbf{2}}$. The coefficient γ is a multiplicative constant that determines the weight of the power efficiency-related term in the total loss, $\mathcal{L}'$. The value of γ directly affects the diffraction

efficiency of the resulting diffractive QPI network design. Specifically, for the diffractive network shown in Fig. 7.2, it was set to be 0. On the other hand, when γ was taken as 0.1, 0.4 and 5.0, the corresponding diffractive QPI network designs achieved 6.31%, 8.17% and 11.05% diffraction efficiency (η), respectively (see Fig. 7.7).

Implementation details of diffractive QPI network training

The deep learning-based diffractive QPI network training was implemented in Python (v3.7.7) and TensorFlow (v1.15.0, Google Inc.). For the gradient-based optimization, we used the Adam optimizer with its momentum parameter β_1 set to 0.5[301]. The learning rate was taken as 0.01 for all the presented diffractive QPI networks. With the batch size equal to 75, we trained all the diffractive networks for 200 epochs, which takes ~40 hours using a computer with a GeForce GTX 1080 Ti GPU (Nvidia Inc.) and Intel® Core ™ i7-8700 Central Processing Unit (CPU, Intel Inc.) with 64 GB of RAM, running Windows 10 operating system (Microsoft). To avoid any aliasing in the representation of the free-space impulse response depicted in Eq. 7.1, the dimensions of the simulation window were taken as 1024×1024.

The PSNR image metric was calculated as follows:

$$PSNR = 20log_{10}\left(\frac{\alpha\pi}{\sqrt{\left|\alpha\pi\boldsymbol{\phi}[q,p] - \boldsymbol{I_{QPI}}[q,p]\right|^2}}\right), \tag{7.9}$$

For SSIM calculations, we used the built-in function in Tensorflow, i.e., tf.image.ssim, where the two inputs were $\alpha\pi\boldsymbol{\phi}[q,p]$ and $\boldsymbol{I_{QPI}}[q,p]$, representing the ground-truth image and the QPI signal synthesized by the diffractive network, respectively. The input parameter "max_val" was set to be $\alpha\pi$ in these SSIM calculations. We should note that for all the images used in our

performance quantification, the SSIM and PSNR metrics were computed over the same output field-of-view, which is approximately 42.4λ×42.4λ.

Chapter 8 Classification and Reconstruction of Spatially Overlapping Phase Images Using Diffractive Optical Networks

Parts of this chapter have previously been published in D. Mengu et al. “Classification and Reconstruction of Spatially Overlapping Phase Images Using Diffractive Optical Networks”, Scientific Reports, DOI: 10.1038/s41598-022-12020-y. This chapter expands upon the previous chapter and uses the coherent signal processing capabilities of diffractive networks to solve phase ambiguity that arises when two phase objects spatially overlap.

Diffractive optical networks unify wave optics and deep learning to all-optically compute a given machine learning or computational imaging task as the light propagates from the input to the output plane. Here, we report the design of diffractive optical networks for the classification and reconstruction of spatially overlapping, phase-encoded objects. When two different phase-only objects spatially overlap, the individual object functions are perturbed since their phase patterns are summed up. The retrieval of the underlying phase images from solely the overlapping phase distribution presents a challenging problem, the solution of which is generally not unique. We show that through a task-specific training process, passive diffractive optical networks composed of successive transmissive layers can all-optically and simultaneously classify two different randomly-selected, spatially overlapping phase images at the input. After trained with ~550 million unique combinations of phase-encoded handwritten digits from the MNIST dataset, our blind testing results reveal that the diffractive optical network achieves an accuracy of >85.8% for all-optical classification of two overlapping phase images of new handwritten digits. In addition to all-optical classification of overlapping phase objects, we also demonstrate the reconstruction of these phase images based on a shallow electronic neural

network that uses the highly compressed output of the diffractive optical network as its input (with e.g., ~20-65 times less number of pixels) to rapidly reconstruct both of the phase images, despite their spatial overlap and related phase ambiguity. The presented phase image classification and reconstruction framework might find applications in e.g., computational imaging, microscopy and quantitative phase imaging fields.

8.1 Introduction

Diffractive Deep Neural Networks (D^2NN)[302] have emerged as an optical machine learning framework that parameterizes a given inference or computational task as a function of the physical traits of a series of engineered surfaces/layers that are connected by diffraction of light. Based on a given task and the associated loss function, deep learning-based optimization is used to configure the transmission or reflection coefficients of the individual pixels/neurons of the diffractive layers so that the desired function is approximated in the optical domain through the light propagation between the input and output planes of the diffractive optical network[76,78–80,103,166,167,169,172,231,302–315]. Upon the convergence of this deep learning-based training phase using a computer, the resulting diffractive surfaces are fabricated using, e.g. 3D printing or lithography, to physically form the diffractive optical network which computes the desired task or inference, without the need for a power source, except for the illumination light.

A diffractive optical network can be considered as a coherent optical processor, where the input information can be encoded in the phase and/or amplitude channels of the sample/object field-of-view. Some of the previous demonstrations of diffractive optical networks utilized 3D printed diffracted layers operating at terahertz (THz) wavelengths to reveal that they can generalize to unseen data achieving >98% and >90% blind testing accuracies for amplitude-encoded handwritten digits (MNIST) and phase-encoded fashion products (Fashion-MNIST), respectively, using passive diffractive layers that collectively compute the all-optical inference at the output plane of the diffractive optical network [79,103,302]. In a recent work[166], diffractive optical networks have been utilized to all-optically infer the data classes of input objects that are illuminated by a broadband light source using only a single-pixel detector at the output plane. This work demonstrated that a broadband diffractive optical network can be trained to extract

and encode the spatial features of input objects into the power spectrum of the diffracted light to all-optically reveal the object classes based on the spectrum of the incident light on a single-pixel detector. Deep learning-based training of diffractive optical networks have also been utilized in solving challenging inverse optical design problems e.g., ultra-short pulse shaping and spatially-controlled wavelength demultiplexing[80,169].

In general, coherent optical processing and the statistical inference capabilities of diffractive optical networks can be exploited to solve various inverse imaging and object classification problems through low-latency, low-power systems composed of passive diffractive layers. One such inverse problem arises when different phase objects reside on top of each other within the sample field-of-view of a coherent imaging platform: the spatial overlap between phase-only thin samples inevitably causes loss of spatial information due to the summation of the overlapping phase distributions describing the individual objects, hence, creating spatial phase ambiguity at the input field-of-view.

Here, we present phase image classification diffractive optical networks that can solve this phase ambiguity problem and simultaneously classify two spatially overlapping images through the same trained diffractive optical network (see Fig. 8.1). In order to address this challenging optical inference problem, we devised four alternative diffractive optical network designs (Figs. 8.1b-e) to all-optically infer the data classes of spatially overlapping phase objects. We numerically demonstrated the efficacy of these diffractive optical network designs in revealing the individual classes of overlapping phase objects using training and testing datasets that are generated based on phase-encoded MNIST digits[202]. Our diffractive optical networks were trained using ~**550** *million* different input phase images containing spatially overlapping

MNIST digits (from the same class as well as different classes); blind testing of one of the resulting diffractive optical networks using 10,000 test images of overlapping phase objects revealed a classification accuracy of >85.8%, optically matching the correct labels of both phase objects that were spatially overlapping within the input field-of-view.

In addition to all-optical classification of overlapping phase images using a diffractive optical network, we also combined our diffractive optical network models with separately trained, electronic image reconstruction networks to recover the individual phase images of the spatially overlapping input objects solely based on the optical class signals collected at the output of the corresponding diffractive optical network. We quantified the success of these digitally reconstructed phase images using the structural similarity index measure (SSIM) and the peak-signal-to-noise-ratio (PSNR) to reveal that a shallow electronic neural network with 2 hidden layers can simultaneously reconstruct both of the phase objects that are spatially overlapping at the input plane despite the fact that the number of detectors/pixels at the output plane of the diffractive optical network is e.g., ~20-65 times smaller compared to an ideal diffraction-limited imaging system. This means the diffractive optical network encoded the spatial features of the overlapping phase objects into a much smaller number of pixels at its output plane, which was successfully decoded by the shallow electronic network to simultaneously perform two tasks: (1) image reconstruction of overlapping spatial features at the input field-of-view, and (2) image decompression.

We believe that the presented diffractive optical network training and design techniques for computational imaging of phase objects will enable memory-efficient, low-power and high frame-rate alternatives to existing phase imaging platforms that often rely on high-pixel count

sensor arrays, and therefore might find applications in e.g. microscopy and quantitative phase imaging fields.

8.2 Results

Spatial overlap between phase objects within the input field-of-view of an optical imaging system obscures the information of samples due to the superposition of the individual phase channels, leading to loss of structural information. For thin phase-only objects (such as e.g., cultured cells or thin tissue sections), when two samples $e^{j\theta_1(x,y)}$ and $e^{j\theta_2(x,y)}$ overlap with each other in space, the resulting object function can be expressed as $e^{j(\theta_1(x,y)+\theta_2(x,y))}$, and therefore a coherent optical imaging system does not have direct access to $\theta_1(x,y)$ or $\theta_2(x,y)$, except their summation (see Fig. 8.1a). In the context of diffractive optical networks and all-optical image classification tasks, another challenging aspect of dealing with spatially overlapping phase objects is that the effective number of data classes represented by different input images significantly increases compared to a single-object classification task. Specifically, for a target dataset with M data classes represented through the phase channel of the input, the total number of data classes at the input (with two overlapping phase objects) becomes $C\binom{M}{2} + M = \frac{M(M-1)}{2} + M$, where C refers to the combination operation . This means that if the diffractive optical network design assigns a single output detector to represent each one of these combinations, one would need $\frac{M(M-1)}{2} + M$ individual detectors. With the use of a differential detection scheme[79] that replaces each class detector with a pair of detectors (virtually representing the positive and negative signals), then the number of detectors at the output plane further increases to $2 \times \left(\frac{M(M-1)}{2} + M\right)$.

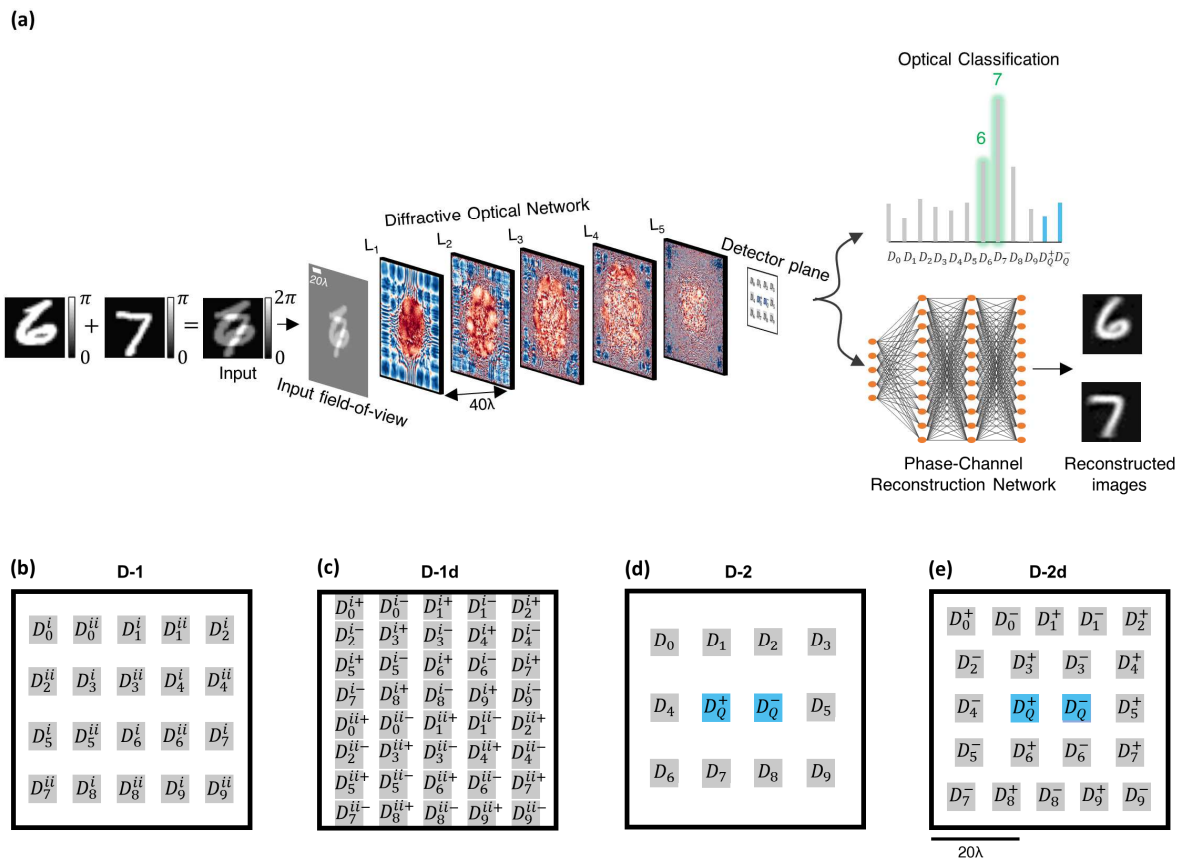

Fig. 8.1 Schematic of a diffractive optical network that can all-optically classify overlapping phase objects despite phase ambiguity at the input; this diffractive optical network also compresses the input spatial information at its output plane for simultaneous reconstruction of the individual phase images of the overlapping input objects using a back-end electronic neural network. (a), Optical layout of the presented 5-layer diffractive optical networks that can all-optically classify overlapping phase objects, e.g., phase-encoded handwritten digits, despite the phase ambiguity at the input plane due to spatial overlap. The diffractive optical network processes the incoming object waves created by the spatially overlapping, phase-encoded digits e.g., '6' and '7', to correctly reveal the classes of both input objects (green). A separately trained shallow electronic neural network (with 2 hidden layers) rapidly reconstructs the individual phase images of both input objects using the optical signals detected at the output plane of the diffractive optical network. (b)-(d), Different detector configurations and class encoding schemes at the output plane of a diffractive optical network, devised to represent all the possible data class combinations at the input field-of-view created by overlapping phase objects.

To mitigate this challenge, in this work we introduced different class encoding schemes that better handle the all-optical classification of these large number of possible class combinations at the input. The output detector layout, D-1, shown in Fig. 8.1b illustrates one alternative design strategy where the problem of classification of overlapping phase objects is solved by using only $2M$ individual detectors with a significant reduction in the number of output detectors when compared to $\frac{M(M-1)}{2} + M$. The use of $2M$ single-pixel detectors at the output plane (see Fig. 8.1b), can handle all the combinations and classify the overlapping input phase objects even if they belong to the same data class or not. To achieve this, we have two different sets of detectors, $\{D_m^i, m = 0,1,2, \dots, M-1\}$ and $\{D_m^{ii}, m = 0,1,2, \dots, M-1\}$, which represent the classes of the individual overlapping phase images. The final class assignments in this scheme are given based on the largest two optical signals among all the $2M$ detectors, where the assigned indices (m) of the corresponding two winner detectors indicate the all-optical classification results for the overlapping phase images. This is a simple class decision rule with a look up table of detector-class assignments (as shown Fig. 8.1b), where the strongest two detector signals indicate the inferred classes based on their m. Stated mathematically, the all-optical estimation of the classes, $\hat{\boldsymbol{c}} = [\widehat{c_1}, \widehat{c_2}]$, of the overlapping phase images is given by,

$$\hat{\boldsymbol{c}} = mod(argmax_2(\boldsymbol{I}), M) \tag{8.1}$$

where $\boldsymbol{I}$ denotes the optical signals detected by $2M$ individual detectors, i.e., $[D_m^i , D_m^{ii}]$. With the $mod(*)$ operation in Eq. 8.1, it can be observed that when the ground truth object classes, c_1 and c_2, are identical, a correct optical inference would result in $\widehat{c_1} = \widehat{c_2}$. On the other hand, when $c_1 \neq c_2$, there are four different detector combinations for the two largest optical signals that would result in the same $(\widehat{c_1}, \widehat{c_2})$ pair according to our class decision rule. For example, in the

case of the input transmittance shown in Fig. 8.1a, which is comprised of handwritten digits '6' and '7', the output object classes based on our decision rule would be the same if the two largest optical signals collected by the detectors correspond to: (1) D_6^i and D_7^{ii}, (2) D_7^i and D_6^{ii}, (3) D_6^i and D_7^i or (4) D_6^{ii} and D_7^{ii}; all of these four combinations of winner detectors at the output plane would reveal the correct classes for the input phase objects in this example (digits '6' and '7').

Therefore, the training the diffractive optical networks according to this class decision rule requires subtle but vital changes in the ground truth labels representing the inputs and the loss function driving the evolution of the diffractive layers compared to a single-object classification system. If we denote the one-hot vector labels representing the classes of the input objects in a single-object classification system as, $\boldsymbol{g^1}$ and $\boldsymbol{g^2}$, with an entry of 1 at their c_1^{th} and c_2^{th} entries, respectively, for the case of spatially overlapping two phase objects at the input field-of-view we can define new ground truth label vectors of length $2M$ using $\boldsymbol{g^1}$ and $\boldsymbol{g^2}$. For the simplest case of $c_1 = c_2$ (i.e., $\boldsymbol{g^1} = \boldsymbol{g^2}$), the $2M$-vector $\boldsymbol{g^e}$ is constructed as $\boldsymbol{g^e} = 0.5 \times [\boldsymbol{g^1}, \boldsymbol{g^2}]$. The constant multiplicative factor of 0.5 ensures that the resulting vector $\boldsymbol{g^e}$ defines a discrete probability density function satisfying $\sum_1^{2M} {g^e}_m = 1$. It is important to note that since $c_1 = c_2$, we have $[\boldsymbol{g^1}, \boldsymbol{g^2}] = [\boldsymbol{g^2}, \boldsymbol{g^1}]$. On the other hand, when the overlapping input phase objects are from different data classes i.e., $c_1 \neq c_2$, we define four different label vectors $\{\boldsymbol{g^a}, \boldsymbol{g^b}, \boldsymbol{g^c}, \boldsymbol{g^d}\}$ representing all the four combinations. Among this set of label vectors, we set $\boldsymbol{g^a} = 0.5 \times [\boldsymbol{g^1}, \boldsymbol{g^2}]$ and $\boldsymbol{g^b} = 0.5 \times [\boldsymbol{g^2}, \boldsymbol{g^1}]$. The label vectors $\boldsymbol{g^c}$ and $\boldsymbol{g^d}$ depict the cases, where the output detectors corresponding to the input object classes lie within D_m^i and D_m^{ii}, respectively. In other words, the c_1^{th} and c_2^{th} entries of $\boldsymbol{g^c}$ are equal to 0.5, and similarly the $(M + c_1)^{th}$ and $(M + c_2)^{th}$ entries of $\boldsymbol{g^d}$ are equal to 0.5, while all the rest of the entries are equal to zero.

Based on these definitions, the training loss function ($\mathcal{L}$) of the associated forward model was selected to reflect all the possible input combinations at the sample field-of-view (input), therefore, it was defined as,

$$\mathcal{L} = (1 - |sgn(c_1 - c_2)|) \times \mathcal{L}_c^e + |sgn(c_1 - c_2)| \times min\{\mathcal{L}_c^a, \mathcal{L}_c^b, \mathcal{L}_c^c, \mathcal{L}_c^d\} \quad (8.2)$$

where $\mathcal{L}_c^a, \mathcal{L}_c^b, \mathcal{L}_c^c, \mathcal{L}_c^d$, and $\mathcal{L}_c^e$ denote the penalty terms computed with respect to the ground truth label vectors $\boldsymbol{g^a}$, $\boldsymbol{g^b}$, $\boldsymbol{g^c}$, $\boldsymbol{g^d}$, and $\boldsymbol{g^e}$, respectively, and $sgn(.)$ is the signum function. The classification errors, $\mathcal{L}_c^x$, are computed using the cross-entropy loss[316],

$$\mathcal{L}_c^x = -\sum_{m=1}^{2M} g_m^x log(\frac{e^{\overline{I_m}}}{\sum_{k=1}^{2M} e^{\overline{I_k}}}) \quad (8.3)$$

where x refers to one of a, b, c, d, or e, $\overline{I_m}$ denotes the normalized intensity collected by a given detector at the output plane (see the Methods section for further details). The term g_m^x in Eq. (8.3) denotes the m^{th} entry of the ground truth data class vector, $\boldsymbol{g^x}$.

Based on this diffractive optical network design scheme and the output detector layout D-1, we trained a 5-layer diffractive optical network (Figs. 8.1,a,b) using the loss function depicted in Eq. 8.2 over ~$550\ million$ input training images containing various combinations of spatially overlapping, phase-encoded MNIST handwritten digits. Following the training phase, the resulting diffractive layers of this network, which we term as D^2NN-D1, are illustrated in Fig. 8.2a. To quantify the generalization performance of D^2NN-D1 for the classification of overlapping phase objects that were never seen by the network before, we created a test dataset, T_2, with 10K phase images, where each image contains two spatially-overlapping phase-encoded test digits randomly selected from the standard MNIST test set, T_1. In this blind testing phase,

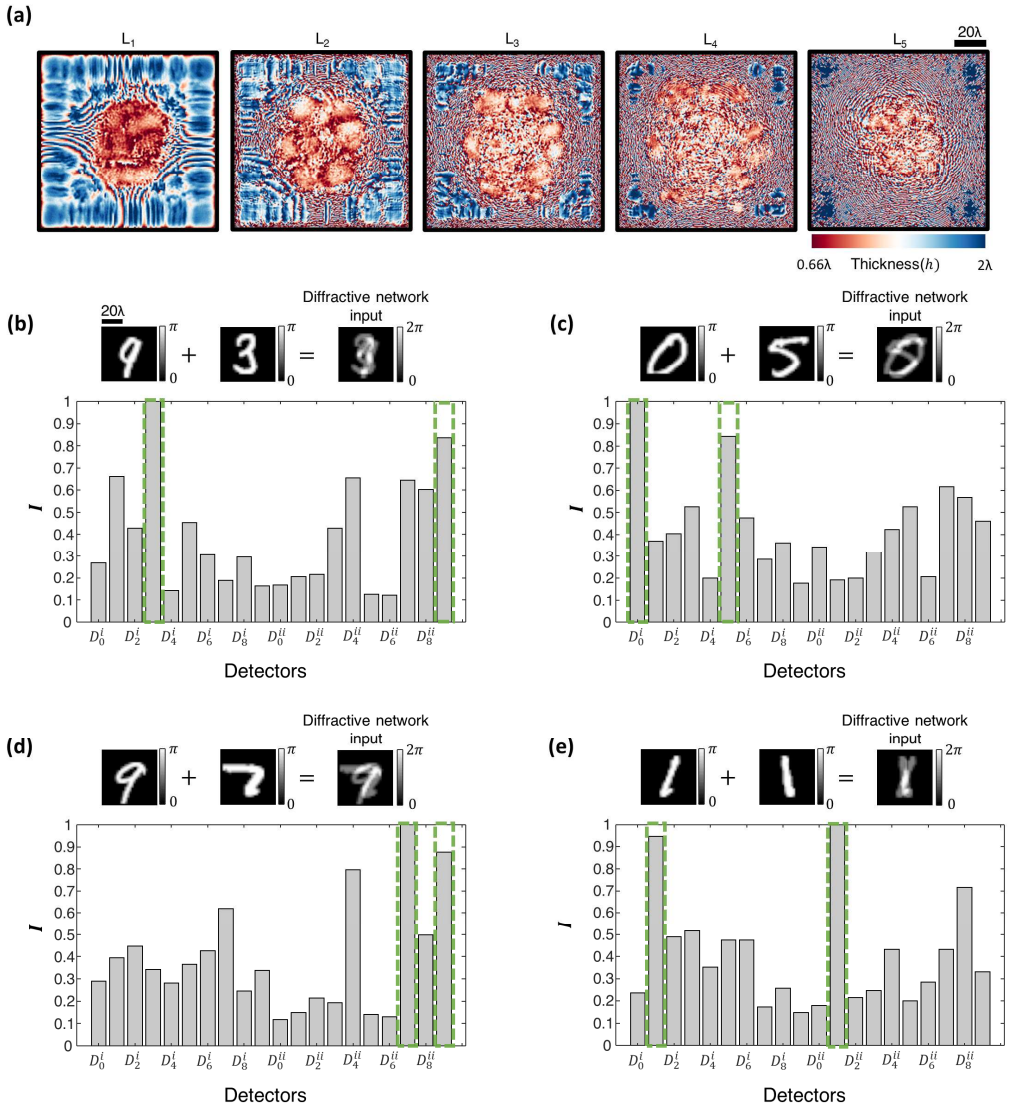

Fig. 8.2 All-optical classification of spatially-overlapping phase objects using the diffractive optical network D2NN-D1, based on the detector layout scheme (D-1). (a) The thickness profiles of the diffractive layers constituting the diffractive optical network D2NN-D1 at the end of its training. This network achieves 82.70% blind inference accuracy on the test image set T2. (b)-(e), Top: Individual phase objects (examples) and the resulting input phase distribution created by their spatial overlap at the input field-of-view. Bottom: The normalized optical signals, I, synthesized by D2NN-D1 at its output detectors. The output detectors with the largest 2 signals correctly reveal the classes of the overlapping input phase objects (indicated with the green rectangular frames).

D^2NN-D1 achieved 82.70% accuracy on T_2, meaning that in 8,270 cases out of 10,000 test inputs, the class estimates $[\widehat{c_1}, \widehat{c_2}]$ at the diffractive optical network's output plane were correct for both of the spatially overlapping handwritten digits. For the remaining 1730 test images, the classification decision of the diffractive optical network is incorrect for at least one of the phase objects within the field-of-view. Figures 8.2b-e depict some of the correctly classified phase image examples from the test dataset T_2 with phase encoded handwritten digits, along with the resulting class scores at the output detectors of the diffractive optical network.

This blind inference accuracy of the diffractive optical network shown in Fig. 8.2a, i.e., D^2NN-D1, can be further improved by combining the above outlined training strategy with a differential detection scheme, where each output detector in D1 (Fig. 8.1b) is replaced with a differential pair of detectors (i.e., a total of 2x2M detectors are located at the output plane, see Fig. 8.1c). The differential signal between a pair of detectors shown in Fig. 8.1c encodes a total of 2xM differential optical signals and similar to the previous approach of D1, the final class assignments in this scheme are given based on the two largest signals among all the differential optical signals. With the incorporation of this differential detection scheme, the vector $\boldsymbol{I}$ in Eq. 8.1 is replaced with the differential signal[79], $\Delta\boldsymbol{I} = \boldsymbol{I}_+ - \boldsymbol{I}_-$, where $\boldsymbol{I}_+$ and $\boldsymbol{I}_-$ denote the optical signals collected by the $2M$ detector pairs, virtually representing the positive and negative parts, respectively.

Using this differential diffractive optical network design, which we termed as D^2NN-D1d (see Fig. 8.1c), we achieved a blind testing accuracy of 85.82% on the test dataset T_2. The diffractive layers comprising the D^2NN-D1d network are shown in Fig. 8.3a, which were trained using $\sim 550\, million$ input phase images of spatially overlapping MNIST handwritten digits,

similar to D^2NN-D1. Compared to the classification accuracy attained by D^2NN-D1, the inference accuracy of its differential counterpart, D^2NN-D1d, is improved by >3.1% at the expense of using $2M$ additional detectors at the output plane of the optical network. Figures 8.3b-e illustrate some examples of the correctly classified phase images from the test dataset T_2 with phase encoded handwritten digits, along with the resulting differential class scores at the output detectors of the diffractive optical network.

The blind inference accuracies achieved by D^2NN-D1 and D^2NN-D1d (82.70% and 85.82%, respectively) on the test dataset T_2, demonstrate the success of the underlying detector layout designs and the associated training strategy for solving the phase ambiguity problem to all-optically classify overlapping phase images using diffractive optical networks. When these two diffractive optical networks (D^2NN-D1 and D^2NN-D1d) are blindly tested over T_1 that provides input images containing a single phase-encoded handwritten digit (without the second overlapping phase object), they attain better classification accuracies of 90.59% and 93.30%, respectively (see the Methods section). As a reference point, a 5-layer diffractive optical network design with an identical layout to the one shown in Fig. 8.1a, can achieve a blind classification accuracy of ~98%[79,103] on test set T_1, provided that it is trained to classify only one phase-encoded handwritten digit per input image (without any spatial overlap with other objects). This reduced classification accuracy of D^2NN-D1 and D^2NN-D1d on test set T_1 (when compared to ~98%) indicates that their forward training model, driven by the loss functions depicted in Eqs. 8.2-8.3, guided the evolution of the corresponding diffractive layers to recognize the spatial features created by the overlapping handwritten digits, as opposed to focusing solely on the actual features describing the individual handwritten digits.

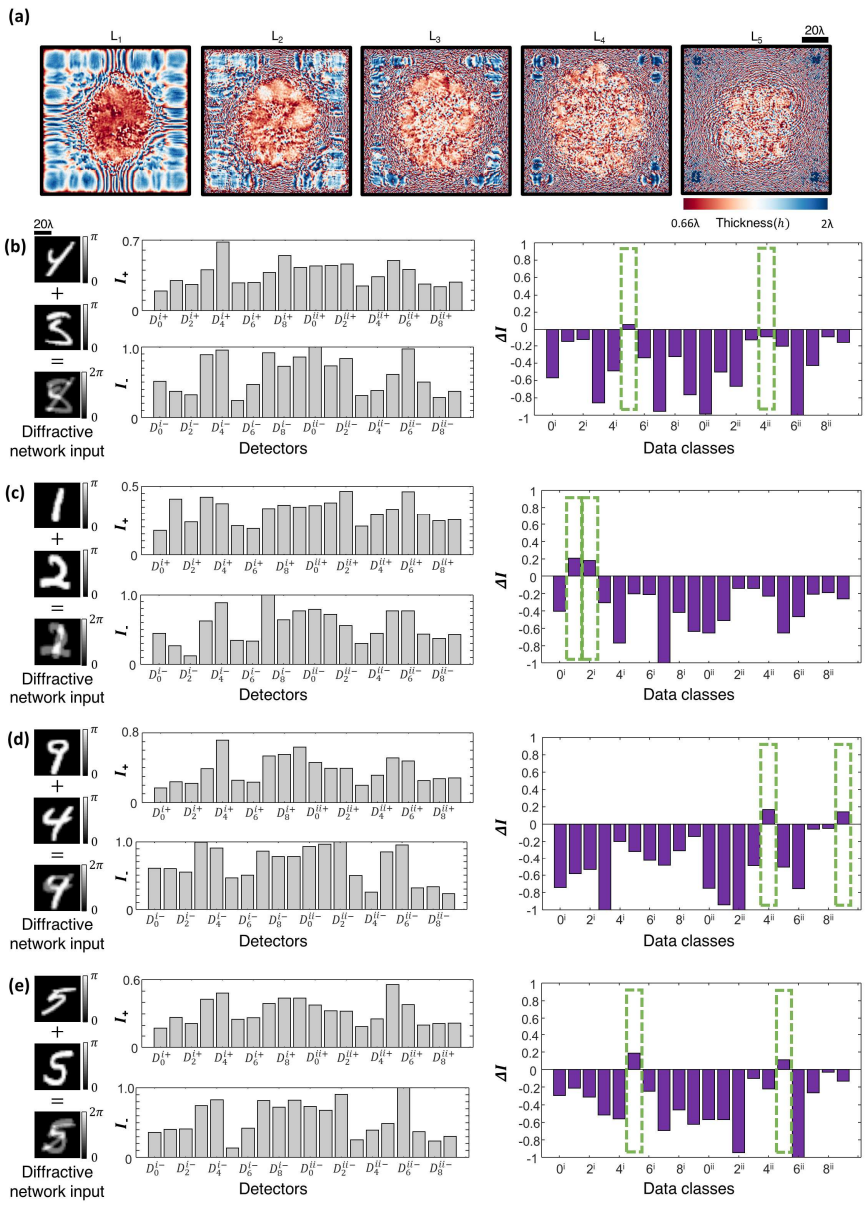

Fig. 8.3 All-optical classification of spatially-overlapping phase objects using the diffractive optical network D2NN-D1d, based on the detector layout scheme D-1d. (a) The thickness profiles of the diffractive layers constituting the diffractive optical network D2NN-D1d at the end of its training. This network achieves 85.82% blind inference accuracy on the test image set T2. (b)-(e), Left: Individual phase objects (examples) and the resulting input phase distribution created by their spatial overlap at the input field-of-view. Middle: The normalized optical signals synthesized by D2NN-D1d at its output detectors. Right: The resulting differential signal. The largest two differential optical signals correctly reveal the classes of the overlapping input phase objects (indicated with the green rectangular frames).

To further reduce the required number of optical detectors at the output plane of a diffractive optical network, we considered an alternative design (D-2) shown in Fig. 8.1d. In this alternative design scheme D-2, there are two extra detectors $\{D_Q^+, D_Q^-\}$ (shown with blue in Fig. 1d), in addition to M class detectors $\{D_m, m = 1,2, \ldots, M\}$ (shown with gray in Fig. 8.1d). The sole function of the additional pair of detectors $\{D_Q^+, D_Q^-\}$ is to decide whether the spatially-overlapping input phase objects belong to the same or different data classes. If the difference signal of this differential detector pair (Fig. 8.1d) is non-negative (i.e., $I_{D_Q^+} \geq I_{D_Q^-}$), the diffractive optical network will infer that the overlapping input objects are from the same data class, hence there is only one class assignment to be made by simply determining the maximum signal at the output class detectors: $\{D_m, m = 1,2, \ldots, M\}$. A negative signal difference between $\{D_Q^+, D_Q^-\}$, on the other hand, indicates that the two overlapping phase objects are from different data classes/digits, and the final class assignments in this case of $I_{D_Q^+} < I_{D_Q^-}$ are given based on the largest two optical signals among all the remaining M detectors at the network output, $\{D_m, m = 1,2, \ldots, M\}$. Refer to the Methods section for further details on the training of diffractive optical networks that employ D-2 (Fig. 8.1d)

Similar to earlier diffractive optical network designs, we used ~$550\ million$ input phase images of spatially overlapping MNIST handwritten digits to train 5 diffractive layers constituting the D^2NN-D2 network (see Fig. 8.4a). Figures 8.4b-d illustrate sample input phase images that contain objects from different data classes, along with the output detector signals that correctly predict the classes/digits of these overlapping phase objects; notice that in each one of these cases, we have at the output plane $I_{D_Q^+} < I_{D_Q^-}$ indicating the success of the network's inference. As another example of blind testing, Fig. 8.4e reports the diffractive optical network's

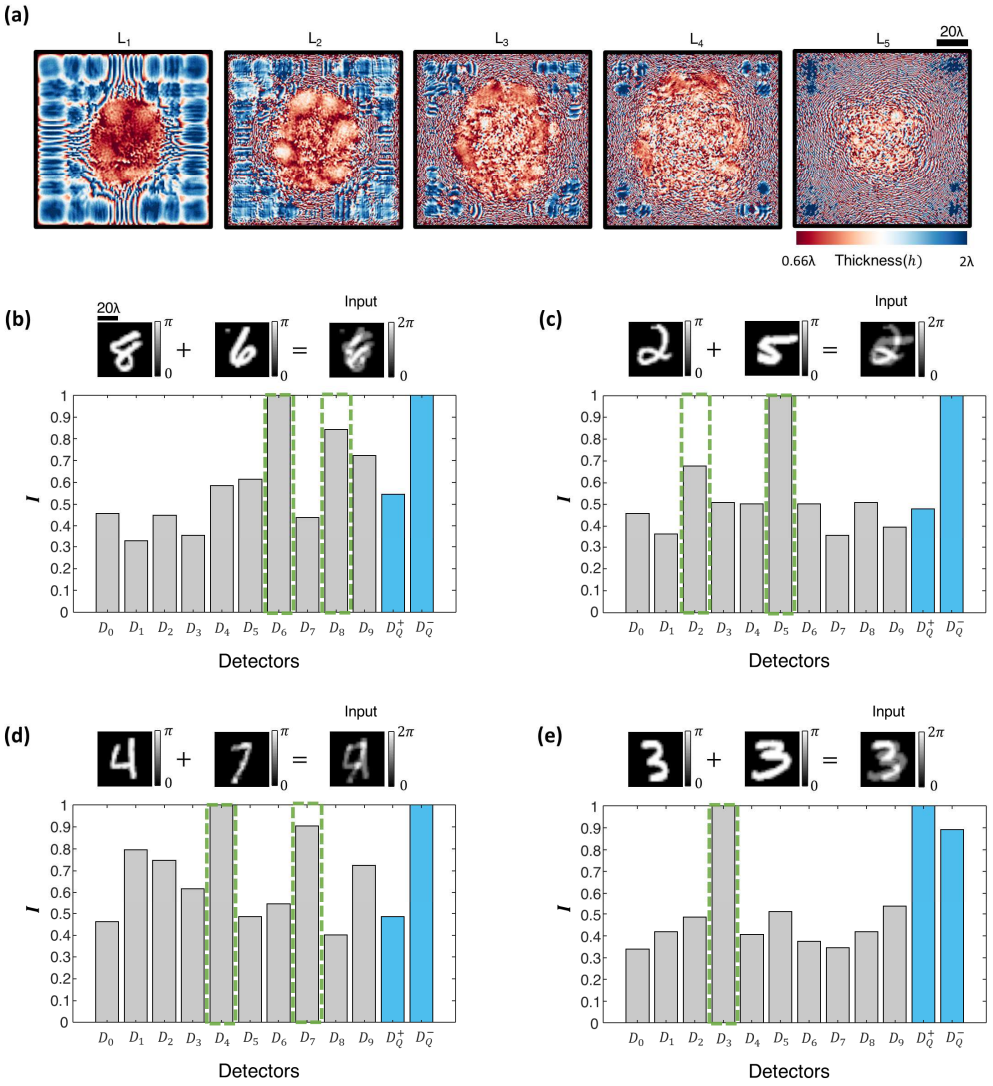

Fig. 8.4 All-optical classification of spatially-overlapping phase objects using the diffractive optical network D2NN-D2, based on the detector layout scheme D-2. (a) The thickness profiles of the diffractive layers constituting the diffractive optical network D2NN-D2 at the end of its training. This network achieves 82.61% blind inference accuracy on the test image set T2. (b)-(e), Top: Individual phase objects (examples) and the resulting input phase distribution created by their spatial overlap at the input field-of-view. Bottom: The normalized optical signals synthesized by D2NN-D2 at its output detectors.

inference for two input phase objects that are from the same data class, i.e., digit '3'. At the network's output, this time we have $I_{D_Q^+} > I_{D_Q^-}$, correctly predicting that the two overlapping phase images are of the same class; the maximum output signal of the remaining output detectors $\{D_m, m = 1,2, \ldots, M\}$ also correctly reveals that the handwritten phase images belong to digit '3' with a maximum signal at D_3. This D^2NN-D2 design provides 82.61% inference accuracy on the test set T_2 with 10K test images, closely matching the inference performance of D^2NN-D1 (82.70%) reported in Fig 8.2. In fact, an advantage of this D^2NN-D2 design lies in its inference performance and blind testing accuracy on test set T_1, achieving 93.38% for classification of input phase images of single digits (without any spatial overlap at the input field-of-view).

We also implemented the differential counterpart of the detector layout D-2, which we term as D-2d (see Fig. 8.1e), where the M class detectors in D-2 are replaced with M differential pairs of output detectors. In this configuration D-2d, the total number of detectors at the output plane of the diffractive optical network becomes $2M + 2$ and the all-optical inference rules remain the same as in D-2: for $I_{D_Q^+} \geq I_{D_Q^-}$, the class inference is made by simply determining the maximum differential signal at the output class detectors, and for the case of $I_{D_Q^+} < I_{D_Q^-}$ the inference of the classes of input phase images is determined based on the largest two differential optical signals at the network output. Figure 8.5a shows the diffractive layers of the resulting D^2NN-D2d that is trained based on the detector layout, D-2d (Fig. 8.1e) using the same training dataset as before: $\sim 550\, million$ input phase images of spatially overlapping, phase-encoded MNIST handwritten digits. This new differential diffractive optical network design, D^2NN-D2d, provides significantly higher blind inference accuracies compared to its non-differential counterpart D^2NN-D2, achieving 85.22% and 94.20% on T_2 and T_1 datasets, respectively.

Figures 8.5b-d demonstrate some examples of the input phase images from test set T_2 that are correctly classified by D^2NN-D2d along with the corresponding optical signals collected by the output detectors representing the positive and negative parts, $\boldsymbol{I_{M+}}$ and $\boldsymbol{I_{M-}}$, of the associated differential class signals, $\boldsymbol{\Delta I_M = I_{M+} - I_{M-}}$. As another example, the input phase image depicted in Fig. 8.5e has two overlapping phase-encoded digits from the same data class, handwritten digit '4', and the diffractive optical network correctly outputs $I_{D_Q^+} > I_{D_Q^-}$ with the maximum differential class score strongly revealing an optical inference of digit '4'.

Table 8.1 summarizes the optical blind classification accuracies achieved by different diffractive optical network designs, D^2NN-D1, D^2NN-D1d, D^2NN-D2 and D^2NN-D2d on test image sets T_2 and T_1. Even though D^2NN-D1d achieves the highest inference accuracy for the classification of spatially overlapping phase objects, D^2NN-D2d offers a balanced optical inference system achieving very good accuracy on both T_1 and T_2. These two differential diffractive optical network models outperform their non-differential counterparts with superior inference performance on both T_2 and T_1.

Next, we aimed to reconstruct the individual images of the overlapping phase objects (handwritten digits) using the detector signals at the output of a diffractive optical network; stated differently our goal here is to resolve the phase ambiguity at the input plane and reconstruct both of the input phase images, despite their spatial overlap and the loss of phase information. For this aim, we combined each one of our diffractive optical networks, D^2NN-D1, D^2NN-D1d, D^2NN-D2 and D^2NN-D2d, one by one, with a shallow, fully-connected (FC) electronic network with two hidden layers, forming a task-specific imaging system as shown in Fig. 8.6. In these hybrid machine vision systems, the optical signals synthesized by a given

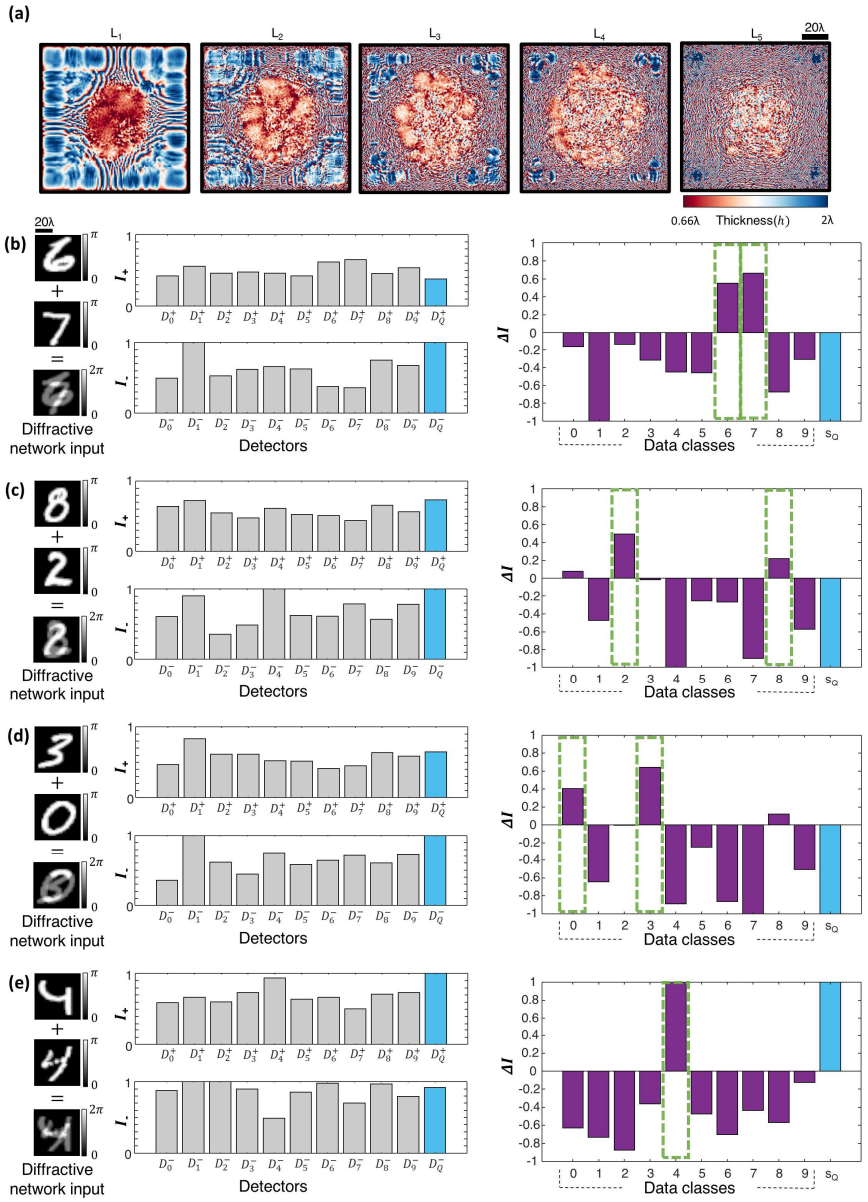

Fig. 8.5 All-optical classification of spatially-overlapping phase objects using the diffractive optical network D2NN-D2d, based on the detector layout scheme D-2d. (a) The thickness profiles of the diffractive layers constituting the diffractive optical network D2NN-D2d at the end of its training. This network achieves 85.22% blind inference accuracy on the test image set T2. (b)-(e), Left: Individual phase objects (examples) and the resulting input phase distribution created by their spatial overlap at the input field-of-view. Middle: The normalized optical signals synthesized by D2NN-D2d at its output detectors. Right: The differential optical signal (purple).

diffractive optical network (front-end encoder) are interpreted as encoded representations of the spatial information content at the input plane. Accordingly, the electronic back-end neural network is trained to process the encoded optical signals collected by the output detectors of the diffractive optical network to decode and reconstruct the individual phase images describing each object function at the input plane, resolving the phase ambiguity due to the spatial overlap of the two phase objects. Figures 8.6a-d illustrate 3 different input images taken from the test set T_2 for each diffractive optical network design (D^2NN-D1, D^2NN-D1d, D^2NN-D2 and D^2NN-D2d) along with the corresponding image reconstructions at the output of each one of the electronic networks that are separately trained to work with the diffractive optical front-end network. As depicted in Fig. 8.6, the electronic image reconstruction networks only have 2 hidden layers with 100 and 400 neurons, and the final output layers of these networks have 28×28×2 neurons, revealing the images of the individual phase objects, resolving the phase ambiguity due to the spatial overlap of the input phase images. The quality of these image reconstructions is quantified using (1) the structural similarity index measure (SSIM) and (2) the peak signal-to-noise ratio (PSNR). Table 8.2 shows the mean SSIM and PSNR values achieved by these hybrid machine vision systems along with the corresponding standard deviations computed over the entire 10K test images (T_2). For these presented image reconstructions, we should emphasize that the dimensionality reduction (i.e., the image data compression) between the input and output planes of the diffractive optical networks (D^2NN-D1, D^2NN-D1d, D^2NN-D2 and D^2NN-D2d) is 39.2×, 19.6×, 65.33× and 35.63×, respectively, meaning that the spatial information of the overlapping phase images at the input field-of-view is significantly compressed (in terms of the number of pixels) at the output plane of the diffractive optical network. This large compression sets another significant challenge for the image reconstruction task in addition to the phase

ambiguity and spatial overlap of the target images. With these large compression ratios, the presented diffractive optical network-based machine vision systems managed to faithfully recover the phase images of each input object despite their spatial overlap and phase information loss, demonstrating the coherent processing power of diffractive optical networks as well as the unique design opportunities enabled by their collaboration with electronic neural networks that form task-specific back-end processors.

8.3 Discussion

The optical classification of overlapping phase images using diffractive optical networks presents a challenging problem due to the spatial overlap of the input images and the associated loss of phase information at the input plane. Interestingly, different combinations of handwritten digits at the input present different amounts of spatial overlap, which is a function of the class of the selected input digits as well as the style of the handwriting of the person. To shed more light on this, we quantified the all-optical blind inference accuracies of the presented diffractive optical networks as a function of the spatial overlap percentage, ξ, at the input field-of-view; see Fig. 8.7. In the first group of examples shown in Fig. 8.7a, the input fields-of-view contain digits from different data classes ($c_1 \neq c_2$) and in the second group of examples shown in Fig. 8.7b, the spatially overlapping objects are from the same data class, $c_1 = c_2$. The input phase images in T_2 exhibit spatial overlap percentages varying between ~20% and ~100%. Figures 8.7c,d illustrate the change in the optical blind inference accuracy of the diffractive optical network, D^2NN-D1, as a function of the spatial overlap percentage, ξ, for the first ($c_1 \neq c_2$) and the second ($c_1 = c_2$) group of test input images, respectively. When $c_1 \neq c_2$ as in Fig. 8.7c, the optical inference accuracy is hindered by the increasing amount of spatial overlap between the two input phase objects, as in this case, the spatial features of the effective input transmittance function

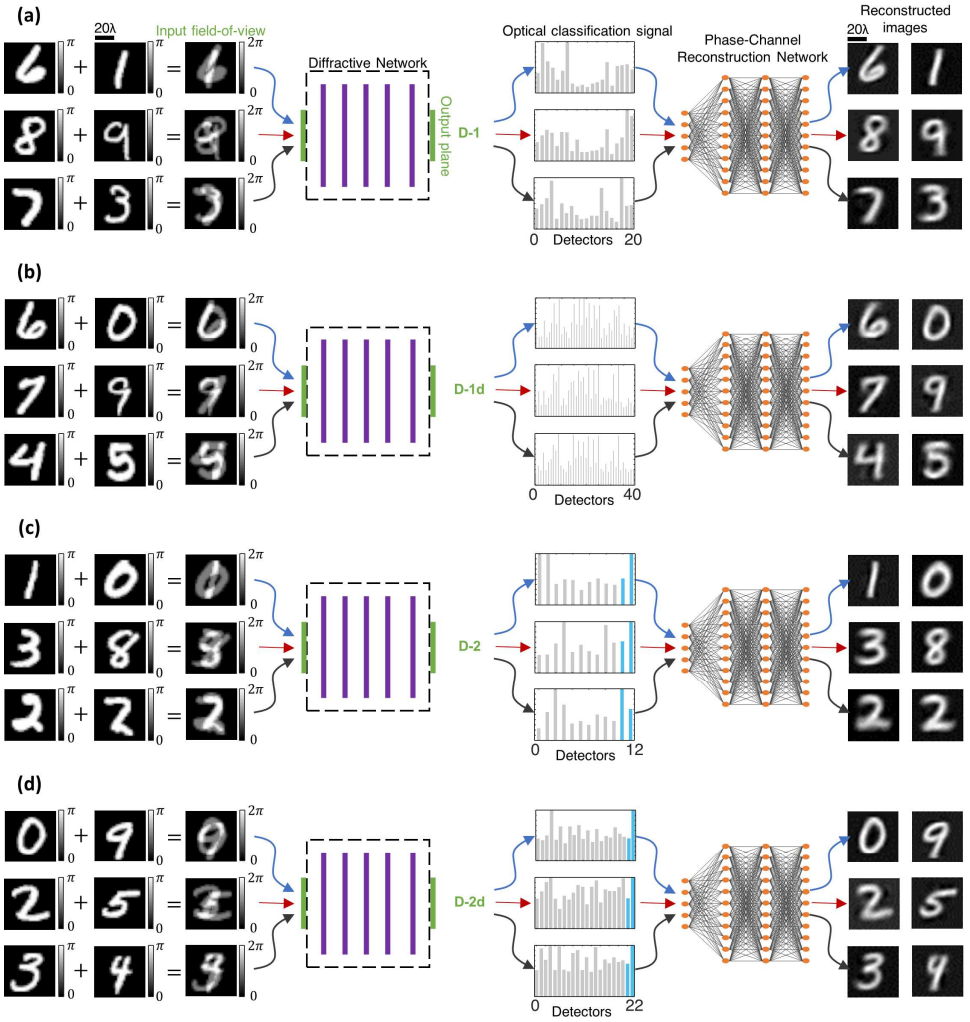

Fig. 8.6 Reconstruction of spatially overlapping phase images using a diffractive optical front-end (encoder) and a separately trained, shallow electronic neural network (decoder) with 2 hidden layers. The front-end diffractive optical networks are (a) D2NN-D1, (b) D2NN-D1d, (c) D2NN-D2, and (d) D2NN-D2d, shown in Figs. 8.2a, 8.3a, 8.4a and 8.5a, respectively. The detector layouts at the output plane of these diffractive optical networks are (a) D-1, (b) D-1d, (c) D-2, and (d) D-2d with 2M, 4M, M+2 and 2M+2 single pixel detectors as shown in Figs. 8.1b-d, respectively; for handwritten digits M=10. These four designs create a compression ratio of 39.2×, 19.6×, 65.33× and 35.63× between the input and output fields-of-view of the corresponding diffractive optical network, respectively. The mean SSIM and PSNR values achieved by these phase image reconstruction networks are depicted in Table 8.2 along with the corresponding standard deviation values computed over the 10K test input images (T2).

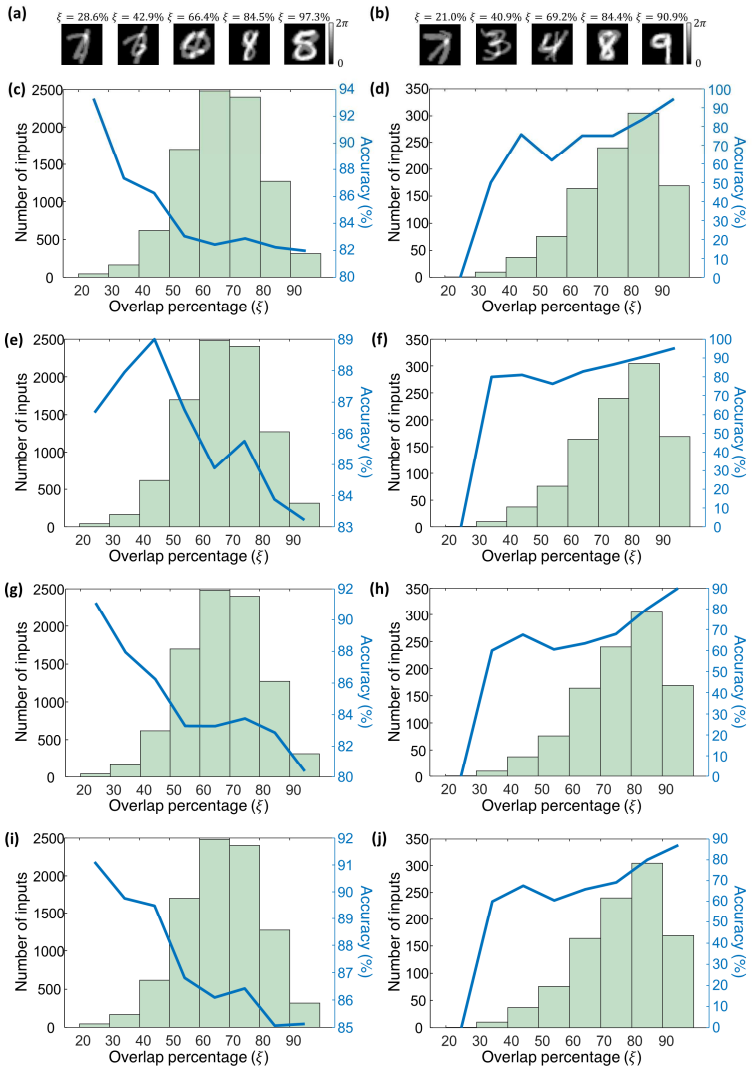

Fig. 8.7 The variation in the optical blind inference accuracies of the presented diffractive optical networks as a function of the spatial overlap percentage (ξ) between the two input phase objects. (a) Sample input images from the test set T2 containing overlapping phase objects from different data classes along with the corresponding overlap percentages, ξ. (b) Same as (a), except the overlapping objects are from the same data class. (c) The blind inference accuracy of the diffractive optical network, D2NN-D1, as a function of the overlap percentage, ξ, and the histogram of ξ, for test inputs in T2 that contain phase objects from two different data classes. (d), Same as (c), except that the test inputs contain phase objects from the same data class. (e) and (f), Same as ((c) and (d)), except, the diffractive optical network design is D2NN-D1d. (g) and (h), Same as ((c) and (d)), except, the diffractive optical network design is D2NN-D2. (i) and (j), Same as ((c) and (d)), except, the diffractive optical network design is D2NN-D2d.

significantly deviate from the features defining the individual data classes. In the other case shown in Fig. 8.7d, i.e., $c_1 = c_2$, the relationship between the spatial overlap ratio ξ and the blind inference accuracy is reversed, since, with $c_1 = c_2$, increasing ξ means that the effective phase distribution at the input plane resembles more closely to a single object/digit. The same behavior can also be observed for the other diffractive optical networks, D^2NN-D1d, D^2NN-D2 and D^2NN-D2d, reported in Figs. 8.7e-f, 8.7g-h, 8.7i-j, respectively.

Next, to verify and test the generalization of the presented diffractive optical network design methods and schemes over different datasets, we trained two new 5-layer diffractive optical networks with detector plane designs identical to the D1d and D2d shown in Figs. 8.1c and 8.1e for the classification of overlapping, phase-encoded objects from a more challenging dataset, Fashion-MNIST. While the D^2NN-D1d shown in Fig. 8.3 can achieve 85.82% accuracy for the classification of overlapping handwritten digits, its equivalent (see Fig. 8.8) that is trained and tested on Fashion-MNIST dataset can attain 73.28% bling testing accuracy. The same comparison also reveals a similar drop in classification accuracy for the D^2NN-D2d model which can classify both of the phase-encoded fashion products in 7221 cases out of 10K blind testing input fields-of-view containing overlapping objects from Fashion-MNIST corresponding to 72.21% accuracy (see Fig. 8.9 for classification examples). On the other hand, considering that a total random guessing would result in an accuracy of 1% for classifying two overlapping objects, these numbers, despite being lower than the case on handwritten digits, demonstrates the efficacy of the presented diffractive optical network design methods. In addition, we combined these two new diffractive optical networks classifying overlapping fashion products with shallow, electronic image reconstruction networks (with 2 hidden layers) forming hybrid vision systems. Despite the lower all-optical classification accuracies, though, the quality of the images

reconstructed by the subsequent electronic networks based solely on the all-optical classification signals synthesized by the associated diffractive optical networks points to an improvement compared to the reconstruction quality achieved for MNIST dataset. For instance, while the hybrid machine vision system with D^2NN-D1d at the optical front-end can provide 0.57±0.10 SSIM score and 16.02±2.21 dB PSNR for the reconstruction of handwritten digits, the hybrid system with the same electronic and optical network architecture achieves 0.61±0.13 and 17.80±2.54 dB for the same metrics, respectively. Similar improvement in the phase channel reconstruction quality also applies for the hybrid systems using D^2NN-D2d as their optical front-end, which, in the case of Fashion-MNIST dataset, attains 0.59±0.13 and 17.18±2.52 dB for SSIM and PSNR, respectively. Exemplary input field-of-view and reconstructed image pairs for both these hybrid systems are depicted in Fig. 8.10.

In summary, to the best of our knowledge, this manuscript reports the first all-optical multi-object classification designs based on diffractive optical networks demonstrating their potential in solving challenging classification and computational imaging tasks in a resource-efficient manner using only a handful detectors at the output plane. In the context of optical classification and reconstruction of overlapping phase objects, also resolving the phase ambiguity due to the spatial overlap of input images, this study exclusively focuses on a setting where the thin input objects are only modulating the phase of the incoming waves, and absorption is negligible. Without loss of generality, the presented diffractive design schemes with the associated loss functions and training methods can also be extended to applications, where the input objects partially absorb the incoming light.

	Test Set	
	T_2 examples	T_1 examples
Diffractive Network	Accuracy (%)	Accuracy (%)
D^2NN-D1	82.70	90.59
D^2NN-D1d	85.82	93.30
D^2NN-D2	82.61	93.38
D^2NN-D2d	85.22	94.20

Table 8.1 The summary of the optical blind inference accuracies achieved by the presented diffractive optical networks on test sets T2 and T1 along with some input examples from these datasets.

		Optical Classification	Image reconstruction	
Diffractive optical network	Number of detectors (M=10)	Accuracy on T_2 (%)	SSIM	PSNR (dB)
D^2NN-D1	$2M$	82.70	0.52±0.12	15.09±2.32
D^2NN-D1d	$4M$	85.82	0.57±0.10	16.02±2.21
D^2NN-D2	M+2	82.61	0.49±0.10	14.55±2.17
D^2NN-D2d	$2M$+2	85.22	0.57±0.12	15.60±2.37

Table 8.2 The comparison of the presented diffractive optical networks, in terms of (1) all-optical overlapping object classification accuracies on T2 and (2) the quality of the image reconstruction achieved through separately-trained, shallow, electronic networks (decoder).

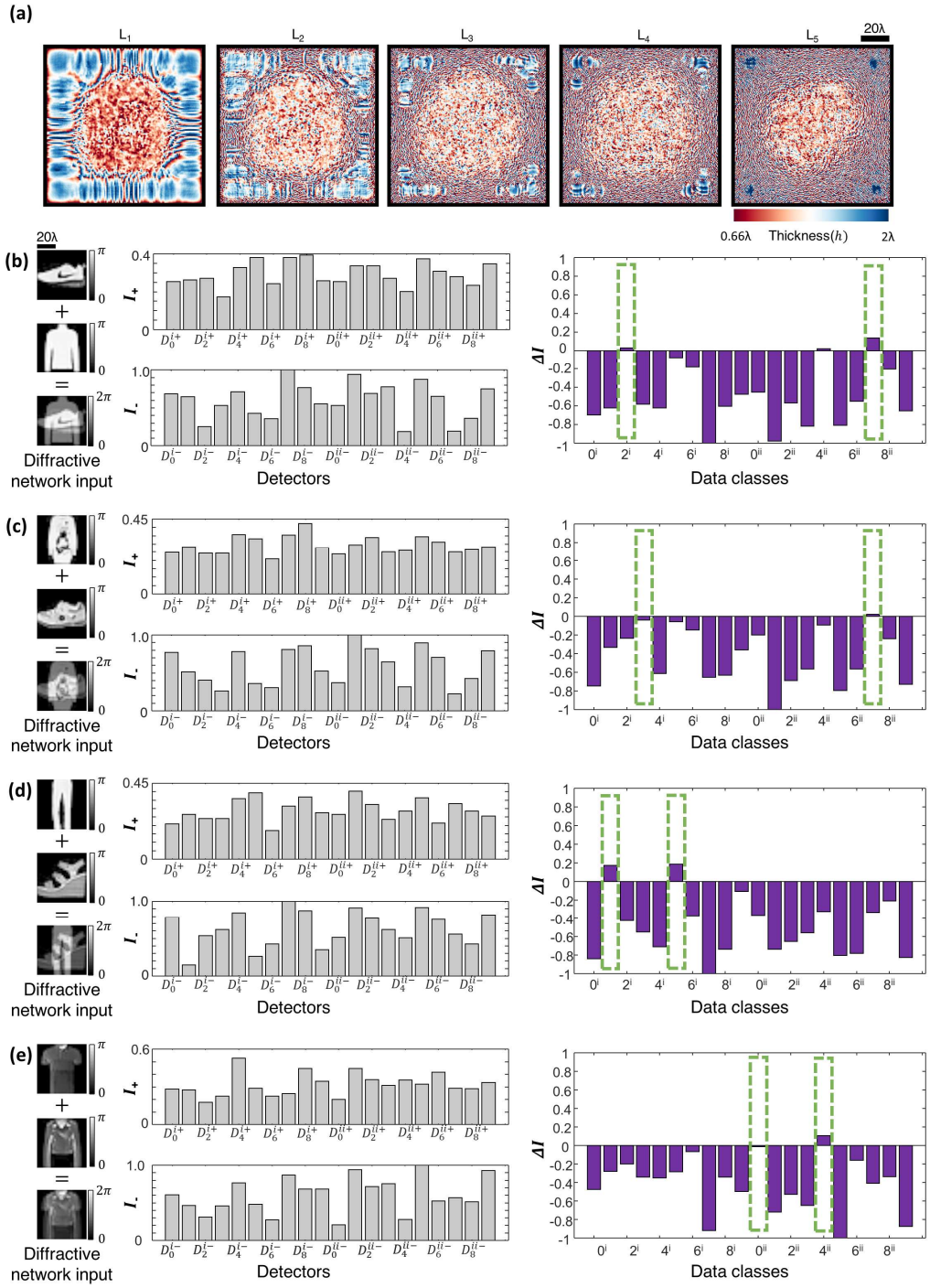

Fig. 8.8 All-optical classification of spatially-overlapping phase objects selected from the Fashion-MNIST dataset (using the D-1d detector layout scheme). (a-e) Same as Fig. 8.3, except that the phase-encoded objects that spatially-overlap within the input field-of-view are randomly selected from the Fashion-MNIST dataset.

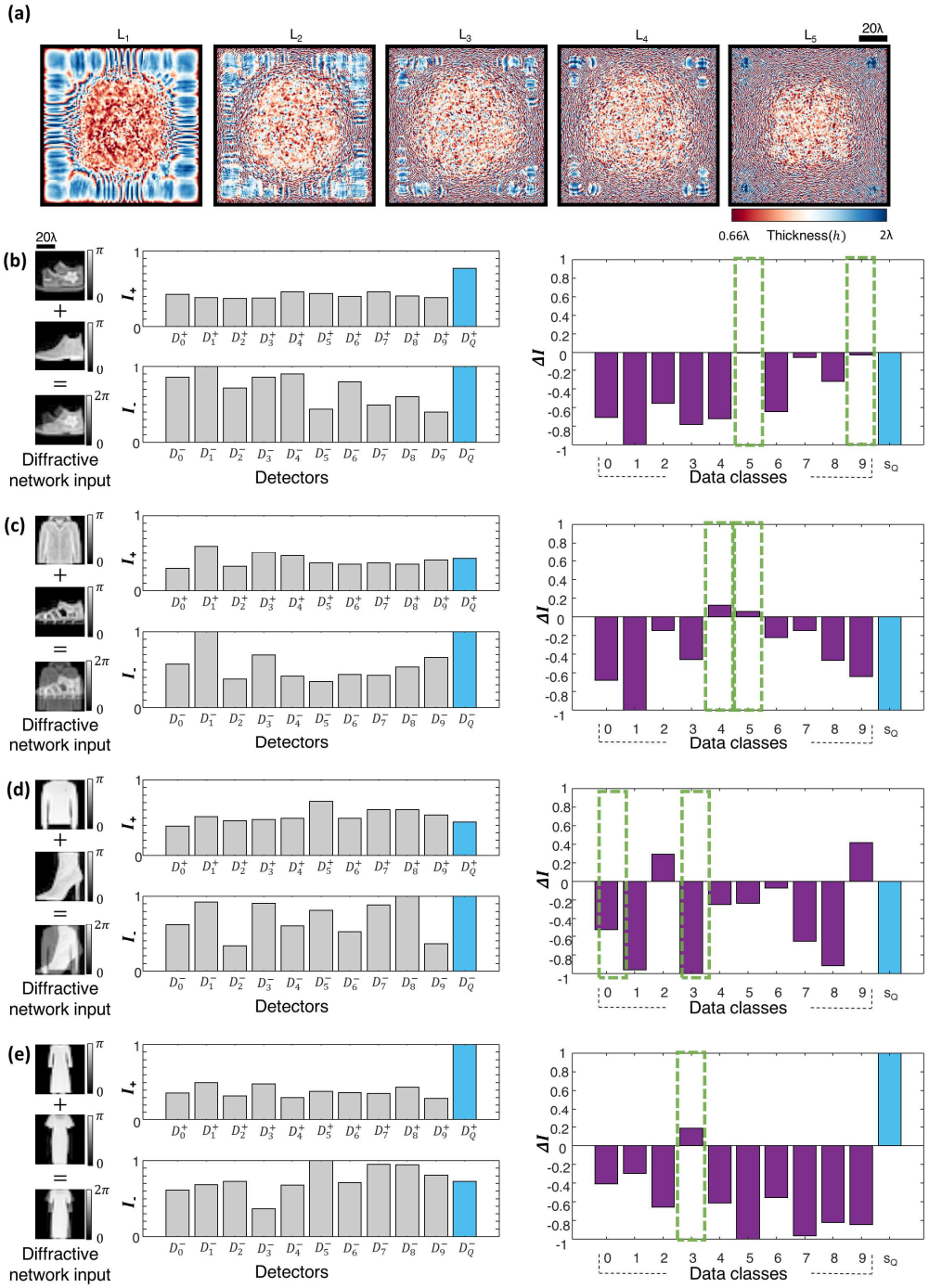

Fig. 8.9 All-optical classification of spatially-overlapping phase objects selected from the Fashion-MNIST dataset (using the D-2d detector layout scheme). (a-e) Same as Fig. 8.5, except that the phase-encoded objects that spatially-overlap within the input field-of-view are randomly selected from the Fashion-MNIST dataset.

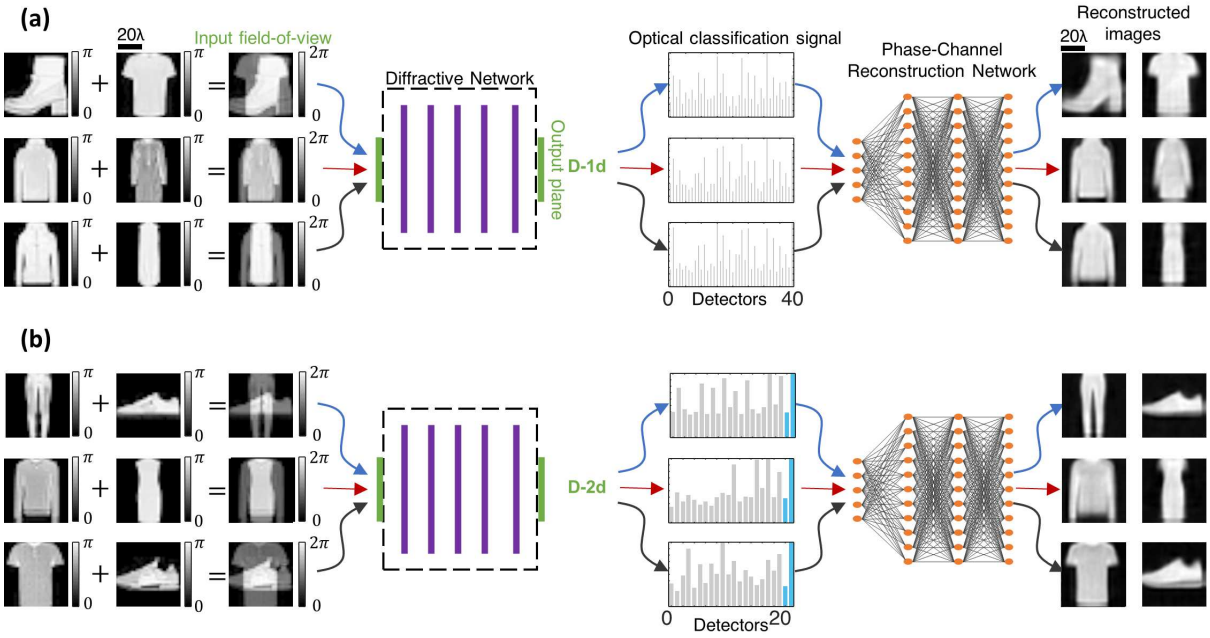

Fig. 8.10 Reconstruction of spatially overlapping phase images using a diffractive optical front-end (encoder) and a separately trained, shallow electronic neural network (decoder) with 2 hidden layers. The front-end diffractive optical networks are (a) D2NN-D1d, and (b) D2NN-D2d shown in Figs. 8.8, and 8.9, respectively. The detector layouts at the output planes of these diffractive optical networks are (a) D-1d and (b) D-2d with 4M and 2M+2 unique detectors as shown in Figs. 8.1b-d, respectively; for fashion products M=10. The mean SSIM and PSNR values achieved by the phase image reconstruction network in (a) are 0.61 and 17.80 dB, in (b) are 0.59 and 17.18 dB, respectively.

8.4 Methods

Design of diffractive optical networks

D^2NN framework formulates a given machine learning e.g., object classification or inverse design task as an optical function approximation problem and parameterizes that function over the physical features of the materials inside a diffractive black-box. As is the case in this study, this optical black-box is often modeled through a series of thin modulation layers connected by the diffraction of light waves. Here, we focused our efforts on 5-layer diffractive optical networks as shown in Fig. 8.1a, each occupying an area of $\mathbf{106\lambda \times 106\lambda}$ on the lateral space with $\boldsymbol{\lambda}$ denoting the wavelength of the illumination light. The modulation function of each diffractive layer was sampled and represented over a 2D regular grid with a period of $0.53\boldsymbol{\lambda}$ resulting in $\boldsymbol{N = 200 \times 200}$ different transmittance coefficients, i.e., the diffractive 'neurons'. Based on the $0.53\boldsymbol{\lambda}$ diffractive feature size, we set the layer-to-layer axial distance to be $40\boldsymbol{\lambda}$ to ensure connectivity between all the neurons on two successive layers.

Following the framework used in the earlier experimental demonstrations of diffractive optical networks based on 3D printed layers[80,103,166], we selected the diffractive layer thickness, $\boldsymbol{h}$, as a trainable physical parameter dictating the transmittance of each neuron together with the refractive index of the material. To limit the material thickness range during the deep learning-based training, $\boldsymbol{h}$ is defined as a function of an auxiliary, learnable variable, $\boldsymbol{h_a}$, and a constant base thickness, $\boldsymbol{h_b}$,

$$\boldsymbol{h = Q_4\left(\frac{\sin(h_a) + 1}{2}(h_m - h_b)\right) + h_b} \tag{8.4},$$

where the function, $\boldsymbol{Q}_*(.)$ represents the *-bit quantization operator and $\boldsymbol{h_m}$ is the maximum allowed material thickness. If the material thickness over the $\boldsymbol{i^{th}}$ diffractive neuron located at $(\boldsymbol{x_i, y_i, z_i})$ is denoted by $\boldsymbol{h(x_i, y_i, z_i)}$, then the resulting transmittance coefficient of that neuron, $\boldsymbol{t(x_i, y_i, z_i)}$, is given by,

$$\boldsymbol{t(x_i, y_i, z_i) = \exp\left(-\frac{2\pi\kappa(\lambda)h(x_i, y_i, z_i)}{\lambda}\right)\exp\left(j(n(\lambda) - n_s)\frac{2\pi h(x_i, y_i, z_i)}{\lambda}\right)} \quad (8.5),$$

where $\boldsymbol{n(\lambda)}$ and $\boldsymbol{\kappa(\lambda)}$ are the real and imaginary parts of complex-valued refractive index of the diffractive material at wavelength, $\boldsymbol{\lambda}$. Following the earlier experimental demonstrations of diffractive optical networks, in this work we set the $\boldsymbol{n(\lambda)}$ and $\boldsymbol{\kappa(\lambda)}$ values to be 1.7227 and 0.031, respectively[103]. The parameter $\boldsymbol{n_s}$ in Eq. 8.5 refers to the refractive index of the medium, surrounding the diffractive layers; without loss of generality, we assumed $\boldsymbol{n_s = 1}$ (air). Based on the outlined material properties, the $\boldsymbol{h_m}$ and $\boldsymbol{h_b}$ in Eq. 8.4 were selected as 2λ and 0.66λ, respectively, ensuring that the phase modulation term in Eq. 8.5, $\boldsymbol{(n(\lambda) - n_s)\frac{2\pi h(x_i, y_i, z_i)}{\lambda}}$, can cover the entire [0-2π] phase modulation range per diffractive feature/neuron.

In this study, the light propagation between diffractive layers of the presented diffractive optical networks was modeled through based on the Rayleigh-Sommerfeld diffraction integral which assumes that the propagating light can be expressed as a scalar field, instead of a vector field. With the subwavelength features on our diffractive surfaces, due to the scalar field assumption, our forward training models do not perform exact calculation of the physically synthesized fields. Exact modelling and computation of the light fields diffracted by subwavelength features requires the utilization of vector diffraction theory[175,176]. On the other hand, successful demonstrations of diffractive optical networks with subwavelength diffractive

neurons have been reported numerous times[80,103,166,231,236,302], hinting that the errors introduced by the limitations of the scalar diffraction theory might be neglected in practical application scenarios in optical computing and machine learning. According to the Rayleigh-Sommerfeld theory of diffraction, a neuron located at $(\boldsymbol{x_i, y_i, z_i})$ can be viewed as the source of a secondary wave,

$$\boldsymbol{w_i(x,y,z) = \frac{z - z_i}{r^2}\left(\frac{1}{2\pi r} + \frac{n_s}{j\lambda}\right)\exp(\frac{j2\pi n_s r}{\lambda})} \tag{8.6}$$

where $\boldsymbol{r}$ denotes the radial distance $\sqrt{(\boldsymbol{x - x_i})^2 + (\boldsymbol{y - y_i})^2 + (\boldsymbol{z - z_i})^2}$. Based on this, the output wave emanating from the $\boldsymbol{i^{th}}$ neuron on the $\boldsymbol{k^{th}}$ layer, $\boldsymbol{u_i^k(x,y,z)}$ can be written as,

$$\boldsymbol{u_i^k(x,y,z) = w_i(x,y,z)t(x_i,y_i,z_i)\sum_{q=1}^{N} u_q^{k-1}(x_i,y_i,z_i)} \tag{8.7}$$

The term $\sum_{q=1}^{N} \boldsymbol{u_q^{k-1}(x_i,y_i,z_i)}$ in Eq. 8.9 represents the wave incident on the $\boldsymbol{i^{th}}$ neuron on the $\boldsymbol{k^{th}}$ layer, generated by the neurons on the previous, $\boldsymbol{(k-1)^{th}}$ diffractive layer.

In this study we also assumed that the transmittance function inside the input field-of-view, $\boldsymbol{T_{in}(x,y)}$, covers an area of $53\boldsymbol{\lambda}\times53\boldsymbol{\lambda}$ and without loss of generality, it is illuminated with a uniform plane wave. At the output plane, the width of each single-pixel detector was set to be $6.36\boldsymbol{\lambda}$ on both x and y directions for all four output detector configurations shown in Figs. 8.1b-d. Based on the outlined optical forward model, the diffractive optical networks process the incoming waves generated by the complex-valued transmittance function, $\boldsymbol{T_{in}(x,y)}$, formed by the overlapping thin phase objects and synthesize a 2D intensity distribution at the output plane for all-optical inference of the classes of the overlapping objects. The optical intensity

distribution within the active area of each output detector is integrated to form elements of the vector, $\boldsymbol{I}$ in Eq. 8.1. The number of elements in this optical signal vector, $\boldsymbol{I}$, is equal to the number of output detectors, thus its length is $\boldsymbol{2M}$, $\boldsymbol{4M}$, $\boldsymbol{M+2}$ and $\boldsymbol{2M+2}$ for D^{2}NN-D1, D^{2}NN-D1d, D^{2}NN-D2 and D^{2}NN-D2d, respectively. As part of our forward training model, $\boldsymbol{I}$ is normalized to form, $\bar{\boldsymbol{I}}$,

$$\bar{\boldsymbol{I}} = \frac{\boldsymbol{I}}{\boldsymbol{c}\,\mathbf{max}\{\boldsymbol{I}\}} \tag{8.8}$$

where the coefficient c in Eq. 8.8 serves as the temperature parameter of the softmax function depicted in Eq. 8.3, and it was empirically set to be 0.1 for training of all the diffractive optical networks. It is important to note that this normalization step in Eq. 8.8 is only used during the training stage, and once the training is finished, the forward inference directly uses the detected intensities to decide on the object classes based on the corresponding decision rules. While the vector $\bar{\boldsymbol{I}}$ is directly used in Eq. 8.3 for the D^{2}NN-D1 network, in the case of D^{2}NN-D1d, $\bar{\boldsymbol{I}}$ is split into two vectors of length $\boldsymbol{2M}$, i.e., $\boldsymbol{I}_+$ and $\boldsymbol{I}_-$, representing the signals collected by the positive and negative detectors, and the associated differential signal is computed as $\boldsymbol{\Delta I = I_+ - I_-}$. Accordingly, during the training of D^{2}NN-D1d, the loss function depicted in Eq. 8.3, were computed using $\boldsymbol{\Delta I}$ instead of $\bar{\boldsymbol{I}}$.

For the diffractive optical network D^{2}NN-D2, the output of the normalization defined in Eq. 8.8 was split into two: $\boldsymbol{I_M}$ and $\boldsymbol{I_Q}$. The first part, $\boldsymbol{I_M}$, represents the optical signals coming from the $\boldsymbol{M}$ class specific detectors in the detector layout D-2 (the gray detectors Fig. 8.1d). The latter, $\boldsymbol{I_Q}$, contains two entries describing the positive and the negative parts of the indicator signals, $\boldsymbol{I_{D_Q^+}}$ and $\boldsymbol{I_{D_Q^-}}$ (see the blue detectors Fig. 8.1d). These two extra detectors, $\{\boldsymbol{D_Q^+}, \boldsymbol{D_Q^-}\}$, form

a differential pair that controls the functional form of the class decision rule based on the sign of the difference between the optical signals collected by these detectors. We accordingly determine the class assignments as follows,

$$\hat{\boldsymbol{c}} = \begin{cases} [\boldsymbol{argmax}(\boldsymbol{I_M}), \boldsymbol{argmax}(\boldsymbol{I_M})], & \boldsymbol{if\ I_{D_Q^+} \geq I_{D_Q^-}} \\ \boldsymbol{argmax_2}(\boldsymbol{I_M}), & \boldsymbol{otherwise.} \end{cases} \tag{8.9}$$

To enable the training of diffractive optical networks according to the class assignment rule in Eq. 8.9, we defined a loss function, $\boldsymbol{\mathcal{L} = \mathcal{L}_Q + \mathcal{L}_c}$, that corresponds to the superposition of two different penalty terms, $\boldsymbol{\mathcal{L}_Q}$ and $\boldsymbol{\mathcal{L}_c}$. Here, $\boldsymbol{\mathcal{L}_Q}$ represents the error computed with respect to the binary ground truth indicator signal, $\boldsymbol{g_Q}$,

$$\boldsymbol{g_Q} = \begin{cases} \mathbf{1}, & \boldsymbol{if\ c_1 = c_2} \\ \mathbf{0}, & \boldsymbol{otherwise} \end{cases}. \tag{8.10}$$

Accordingly, $\boldsymbol{\mathcal{L}_Q}$ was defined as,

$$\boldsymbol{\mathcal{L}_Q = -g_Q \log\left(\sigma\left(I_{D_Q^+} - I_{D_Q^-}\right)\right) - (1 - g_Q)\log\left(1 - \sigma\left(I_{D_Q^+} - I_{D_Q^-}\right)\right)} \tag{8.11}$$

where $\boldsymbol{\sigma}(\cdot)$ denotes the sigmoid function. The classification loss, $\boldsymbol{\mathcal{L}_c}$, on the other hand, is identical to the cross-entropy loss depicted in Eq. 8.3, except that the vector $\boldsymbol{I}$ is replaced with $\boldsymbol{I_M}$. Unlike the previous diffractive optical networks (D^2NN-D1 and D^2NN-D1d), the forward model of the diffractive optical networks trained based on the output detector layout D-2 do *not* require multiple ground truth vector labels. Simply the ground truth label vector of a given input field-of-view is defined as $\boldsymbol{g = \frac{g^1 + g^2}{2}}$ satisfying the condition, $\boldsymbol{\sum_1^M g_m = 1}$.

In the case of D^{2}NN-D2d, the vector $\bar{\boldsymbol{I}}$ contains 3 main parts, $\boldsymbol{I_{M+}}$, $\boldsymbol{I_{M-}}$ and $\boldsymbol{I_Q}$ where $\boldsymbol{I_{M+}}$ and $\boldsymbol{I_{M-}}$ are length $\boldsymbol{M}$ vectors containing the normalized optical signals collected by the detectors representing the positive and negative parts of the final differential class scores $\boldsymbol{\Delta I_M = I_{M+} - I_{M-}}$. Accordingly, in the associated forward training model, the intensity vector $\boldsymbol{I_M}$ in Eq. 8.9 is replaced with the differential signal, $\boldsymbol{\Delta I_M}$.

Testing of diffractive optical networks on dataset $\boldsymbol{T_1}$

During the blind testing of the presented diffractive optical networks on the test set T_1, the class estimation solely uses the $\boldsymbol{argmax}$ operation, searching for the highest class-score synthesized by the diffractive optical networks, based on the associated output plane detector layouts shown in Fig. 8.1. The purpose of this performance quantification using T_1 is to reveal whether the diffractive optical networks trained based on overlapping input phase objects can learn and automatically recognize the characteristic spatial features of the individual handwritten digits (without any spatial overlap). For this goal, in the case of D^{2}NN-D1 and D^{2}NN-D1d, the class estimation rule in Eq. 8.1 was replaced with, $\boldsymbol{mod(argmax(I), M)}$ and $\boldsymbol{mod(argmax(\Delta I), M)}$, respectively. Since the input images in the test set T_1 contain single, phase-encoded handwritten digits without the second overlapping phase object, the optical signals collected by the detectors, $\{\boldsymbol{D_Q^+, D_Q^-}\}$, at the output plane of the diffractive optical networks D^{2}NN-D2 and D^{2}NN-D2d become irrelevant for the classification of the images in T_1. Therefore, the decision rule in Eq. 8.9 is simplified to $\boldsymbol{argmax(I_M)}$ and $\boldsymbol{argmax(\Delta I_M)}$ for the all-optical classification of the input test images in T_1 based on the diffractive optical networks D^{2}NN-D2 and D^{2}NN-D2d, respectively.

Architecture and Training of the Phase Image Reconstruction Network

The phase image reconstruction electronic neural networks (back-end) following each of the presented diffractive optical networks (front-end) include 4 neural layers. The number of neurons on their first layer is equal to the number of detectors at the output plane of the preceding diffractive optical network (D-1, D-1d, D-2 or D-2d, see Figs. 8.1a-d). The number of neurons, on the subsequent 3 layers are 100, 400, 1568, respectively. Note that the output layer of each image reconstruction electronic neural network has 2×28×28 neurons as it simultaneously reconstructs both of the overlapping phase objects, resolving the phase ambiguity due to the spatial overlap at the input plane. Each fully-connected (FC) layer constituting these image reconstruction networks applies the following operations:

$$\boldsymbol{\rho_{l+1}} = \mathbf{BN}\{\mathbf{LReLU}[\mathbf{FC}\{\boldsymbol{\rho_l}\}]\} \tag{8.12}$$

where $\boldsymbol{\rho_{l+1}}$ and $\boldsymbol{\rho_l}$ denote the output and input values of the $\boldsymbol{l^{th}}$ layer of the electronic network, respectively. The operator LReLU stands for the leaky rectified linear unit:

$$\boldsymbol{LReLU}[\boldsymbol{x}] = \begin{Bmatrix} \boldsymbol{x}, & \boldsymbol{if\ x \geq 0} \\ \mathbf{0.1}\boldsymbol{x}, & \boldsymbol{otherwise} \end{Bmatrix}. \tag{8.13}$$

The batch normalization, BN, normalizes the activations at the output of LReLU to zero mean and a standard deviation of 1, and then shifts the mean to a new center, $\boldsymbol{\beta^{(l)}}$, and re-scales it with a multiplicative factor, $\boldsymbol{\gamma^{(l)}}$, where $\boldsymbol{\beta^{(l)}}$ and $\boldsymbol{\gamma^{(l)}}$ are learnable parameters, i.e.,

$$\mathrm{BN}[\boldsymbol{x}] = \boldsymbol{\gamma^{(l)}} \cdot \frac{\boldsymbol{x} - \boldsymbol{\mu_B^{(l)}}}{\sqrt{\boldsymbol{\sigma_B^{(l)^2}} + \boldsymbol{\epsilon}}} + \boldsymbol{\beta^{(l)}} \tag{8.14}$$

$$\boldsymbol{\mu_B} = \frac{\mathbf{1}}{\boldsymbol{m}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{m}} \boldsymbol{x_i}, \quad \boldsymbol{\sigma_B^2} = \frac{\mathbf{1}}{\boldsymbol{m}} \sum_{\boldsymbol{i=1}}^{\boldsymbol{m}} (\boldsymbol{x_i} - \boldsymbol{\mu_B})^2$$

The hyperparameter $\boldsymbol{\epsilon}$ is a small constant that avoids division by 0 and it was taken as $\mathbf{10^{-3}}$.

The training of the phase image reconstruction networks was driven by the reversed Huber (or "BerHu") loss, which computes the error between two images, $\boldsymbol{a(x,y)}$ and $\boldsymbol{b(x,y)}$, as follows:

$$\boldsymbol{BerHu(a,b)} = \sum_{\substack{x,y \\ |a(x,y)-b(x,y)|\leq\varphi}} |\boldsymbol{a(x,y)} - \boldsymbol{b(x,y)}| + \sum_{\substack{x,y \\ |a(x,y)-b(x,y)|>\varphi}} \frac{[\boldsymbol{a(x,y)} - \boldsymbol{b(x,y)}]^2 + \boldsymbol{\varphi^2}}{\mathbf{2}\boldsymbol{\varphi}} \tag{8.15}$$

The hyperparameter $\boldsymbol{\varphi}$ in Eq. 8.15 is a threshold for the transition between mean-absolute-error and mean-squared-error, and it was set to be 20% of the standard deviation of the ground truth image.

If we let $\boldsymbol{\phi_p(x,y)}$ and $\boldsymbol{\phi_q(x,y)}$ denote the first and second output of each image reconstruction electronic network (i.e., 28x28 pixels per phase object), the image reconstruction loss, $\boldsymbol{\mathcal{L}_r}$, was defined as the *minimum* of two error terms, $\boldsymbol{\mathcal{L}_r'}$ and $\boldsymbol{\mathcal{L}_r''}$, i.e.,

$$\boldsymbol{\mathcal{L}_r} = \mathbf{min}\{\boldsymbol{\mathcal{L}_r'}, \boldsymbol{\mathcal{L}_r''}\},$$

$$\boldsymbol{\mathcal{L}_r'} = \frac{\left[\mathbf{BerHu}\left(\boldsymbol{\phi_p(x,y)}, \boldsymbol{\phi_1(x,y)}\right) + \mathbf{BerHu}\left(\boldsymbol{\phi_q(x,y)}, \boldsymbol{\phi_2(x,y)}\right)\right]}{\mathbf{2}},$$

$$\boldsymbol{\mathcal{L}_r''} = \frac{\left[\mathbf{BerHu}\left(\boldsymbol{\phi_p(x,y)}, \boldsymbol{\phi_2(x,y)}\right) + \mathbf{BerHu}\left(\boldsymbol{\phi_q(x,y)}, \boldsymbol{\phi_1(x,y)}\right)\right]}{\mathbf{2}}, \tag{8.16}$$

where $\boldsymbol{\phi_1(x,y)}$ and $\boldsymbol{\phi_2(x,y)}$ denote the ground truth phase images of the first and second objects, respectively, which overlap at the input plane of the diffractive optical network. As there is no hierarchy or priority difference between the input objects $\boldsymbol{\phi_1(x,y)}$ and $\boldsymbol{\phi_2(x,y)}$, Eq. 8.16 lets the image reconstruction network to choose their order regarding its output activations.

Miscellaneous details of diffractive network training

With the 0.53λ lateral sampling rate in our forward optical model, the transmittance function inside the field-of-view, $T_{in}(x,y)$, was represented as a 100×100 discrete signal. In our diffractive optical network training, the 8-bit grayscale values of the MNIST digits were first converted to 32-bit double format, normalized to the range [0,1] and then resized to 100×100 using bilinear interpolation. If we denote these normalized and resized grayscale values of the two input objects/digits that overlap at the input plane as $\theta_1(x,y)$ and $\theta_2(x,y)$ then the transmittance function within the input field-of-view, $T_{in}(x,y)$, is defined as,

$$T_{in}(x,y) = e^{j\pi\theta_1(x,y)} e^{j\pi\theta_2(x,y)} \quad (8.17).$$

During the training of the presented diffractive optical networks, $\theta_1(x,y)$ and $\theta_2(x,y)$ are randomly selected from the standard 55K training samples of MNIST dataset without replacement, meaning that, an already selected training digit was not selected again until all 55K samples are depleted constituting an epoch of the training phase. In this manner, we trained the diffractive optical networks for 20,000 epochs, showing each optical network approximately *550 million* different $T_{in}(x,y)$ during the training phase. To generate the input fields in test dataset T_2, we randomly selected $\theta_1(x,y)$ and $\theta_2(x,y)$ among the standard 10K test samples of MNIST dataset, without replacement, and this was repeated two times providing us the 10K unique phase

input images of overlapping handwritten digits constituting T_2. In our T_2 test set, 8998 inputs contain overlapping digits from different data classes, while the remaining 1002 inputs have overlapping samples from the same data class/digit. The validation image set, on the other hand, contains 5K unique phase input images created by randomly selecting $\theta_1(x,y)$ and $\theta_2(x,y)$ among the standard 10K test samples of MNIST dataset without replacement.

The overlap percentage, ξ, between any given pair of samples, $\theta_1(x,y)$ and $\theta_2(x,y)$ (see Fig. 8.7), is quantified by,

$$\xi_1 = \frac{\sum_q \sum_p \left| sgn\left(\theta_2(x_q, y_p)\right)\right| \theta_1(x_q, y_p)}{\sum_{q'} \sum_{p'} \theta_1(x_{q'}, y_{p'})} \times 100$$

$$\xi_2 = \frac{\sum_q \sum_p \left| sgn\left(\theta_1(x_q, y_p)\right)\right| \theta_2(x_q, y_p)}{\sum_{q'} \sum_{p'} \theta_2(x_{q'}, y_{p'})} \times 100$$

$$\xi = \max\{\xi_1, \xi_2\} \tag{8.18}$$

In Equation 8.18, ξ_1 and ξ_2 quantify the percentage of the input pixels that contain the spatial overlap with respect to $\theta_1(x,y)$ and $\theta_2(x,y)$, respectively, and the final ξ is taken as the *max* of these two values.

For the digital implementation of the diffractive optical network training outlined above, we developed a custom-written code in Python (v3.6.5) and TensorFlow (v1.15.0, Google Inc.). The backpropagation updates were calculated using the Adam[106] optimizer with its parameters set to be the default values as defined by TensorFlow and kept identical in each model. The learning rate was set to be 0.001 for all the diffractive optical network models presented here. The training batch size was taken as 75 during the deep learning-based training of the presented

diffractive optical networks. The training of a 5-layer diffractive optical network with 40K diffractive neurons per layer for 20,000 epochs takes approximately 24 days using a computer with a GeForce GTX 1080 Ti Graphical Processing Unit (GPU, Nvidia Inc.) and Intel® Core ™ i7-8700 Central Processing Unit (CPU, Intel Inc.) with 64 GB of RAM, running Windows 10 operating system (Microsoft).

References

1 Caulfield HJ, Kinser J, Rogers SK. Optical neural networks. *Proceedings of the IEEE* 1989; **77**: 1573–1583.

2 Yu FTS, Jutamulia S. *Optical Signal Processing, Computing, and Neural Networks*. 1st ed. John Wiley & Sons, Inc.: New York, NY, USA, 1992.

3 Yu FTS. II Optical Neural Networks: Architecture, Design and Models. In: Wolf E (ed). *Progress in Optics*. Elsevier, 1993, pp 61–144.

4 Psaltis D, Farhat N. Optical information processing based on an associative-memory model of neural nets with thresholding and feedback. *Opt Lett, OL* 1985; **10**: 98–100.

5 Farhat NH, Psaltis D, Prata A, Paek E. Optical implementation of the Hopfield model. *Appl Opt, AO* 1985; **24**: 1469–1475.

6 Wagner K, Psaltis D. Multilayer optical learning networks. *Appl Opt, AO* 1987; **26**: 5061–5076.

7 Psaltis D, Brady D, Wagner K. Adaptive optical networks using photorefractive crystals. *Appl Opt, AO* 1988; **27**: 1752–1759.

8 Psaltis D, Brady D, Gu X-G, Lin S. Holography in artificial neural networks. *Nature* 1990; **343**: 325–330.

9 Weverka RT, Wagner K, Saffman M. Fully interconnected, two-dimensional neural arrays using wavelength-multiplexed volume holograms. *Opt Lett, OL* 1991; **16**: 826–828.

10 Javidi B, Li J, Tang Q. Optical implementation of neural networks for face recognition by the use of nonlinear joint transform correlators. *Appl Opt, AO* 1995; **34**: 3950–3962.

11 Shen Y, Harris NC, Skirlo S, Prabhu M, Baehr-Jones T, Hochberg M *et al.* Deep learning with coherent nanophotonic circuits. *Nature Photonics* 2017; **11**: 441–446.

12 Bueno J, Maktoobi S, Froehly L, Fischer I, Jacquot M, Larger L *et al.* Reinforcement learning in a large-scale photonic recurrent neural network. *Optica, OPTICA* 2018; **5**: 756–760.

13 Hughes TW, Minkov M, Shi Y, Fan S. Training of photonic neural networks through in situ backpropagation and gradient measurement. *Optica, OPTICA* 2018; **5**: 864–871.

14 Shastri BJ, Tait AN, de Lima TF, Nahmias MA, Peng H-T, Prucnal PR. Principles of Neuromorphic Photonics. *arXiv:180100016 [physics]* 2018; : 1–37.

15 Lin X, Rivenson Y, Yardimci NT, Veli M, Luo Y, Jarrahi M *et al.* All-optical machine learning using diffractive deep neural networks. *Science* 2018; **361**: 1004–1008.

16 Chang J, Sitzmann V, Dun X, Heidrich W, Wetzstein G. Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification. *Scientific Reports* 2018; **8**: 12324.

17 Soures N, Steidle J, Preble S, Kudithipudi D. Neuro-MMI: A Hybrid Photonic-Electronic Machine Learning Platform. In: *2018 IEEE Photonics Society Summer Topical Meeting Series (SUM)*. 2018, pp 187–188.

18 Chakraborty I, Saha G, Sengupta A, Roy K. Toward Fast Neural Computing using All-Photonic Phase Change Spiking Neurons. *Scientific Reports* 2018; **8**: 12980.

19 Bagherian H, Skirlo S, Shen Y, Meng H, Ceperic V, Soljacic M. On-Chip Optical Convolutional Neural Networks. *arXiv:180803303 [cs]* 2018.http://arxiv.org/abs/1808.03303 (accessed 16 Sep2018).

20 Tait AN, de Lima TF, Zhou E, Wu AX, Nahmias MA, Shastri BJ *et al.* Neuromorphic photonic networks using silicon photonic weight banks. *Scientific Reports* 2017; **7**. doi:10.1038/s41598-017-07754-z.

21 Mehrabian A, Al-Kabani Y, Sorger VJ, El-Ghazawi T. PCNNA: A Photonic Convolutional Neural Network Accelerator. *arXiv:180708792 [cs, eess]* 2018.http://arxiv.org/abs/1807.08792 (accessed 2 Oct2018).

22 George J, Amin R, Mehrabian A, Khurgin J, El-Ghazawi T, Prucnal PR *et al.* Electrooptic Nonlinear Activation Functions for Vector Matrix Multiplications in Optical Neural Networks. In: *Advanced Photonics 2018 (BGPP, IPR, NP, NOMA, Sensors, Networks, SPPCom, SOF)*. OSA: Zurich, 2018, p SpW4G.3.

23 LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015; **521**: 436–444.

24 Schmidhuber J. Deep learning in neural networks: An overview. *Neural Networks* 2015; **61**: 85–117.

25 Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 1998; **86**: 2278–2324.

26 Xiao H, Rasul K, Vollgraf R. Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms. *arXiv:170807747 [cs, stat]* 2017.http://arxiv.org/abs/1708.07747 (accessed 19 Sep2018).

27 Maas AL, Hannun AY, Ng AY. Rectifier nonlinearities improve neural network acoustic models. In: *Proc. icml*. 2013, p 3.

28 Hunger R. An Introduction to Complex Differentials and Complex Differentiability. ; : 20.

29 Wan L, Zeiler M, Zhang S, Le Cun Y, Fergus R. Regularization of neural networks using dropconnect. In: *International Conference on Machine Learning*. 2013, pp 1058–1066.

30 Golik P, Doetsch P, Ney H. Cross-entropy vs. squared error training: a theoretical and experimental comparison. In: *Interspeech*. 2013, pp 1756–1760.

31 He K, Zhang X, Ren S, Sun J. Identity Mappings in Deep Residual Networks. In: Leibe B, Matas J, Sebe N, Welling M (eds). *Computer Vision – ECCV 2016*. Springer International Publishing, 2016, pp 630–645.

32 Emons M, Obata K, Binhammer T, Ovsianikov A, Chichkov BN, Morgner U. Two-photon polymerization technique with sub-50 nm resolution by sub-10 fs laser pulses. *Optical Materials Express* 2012; **2**: 942.

33 Build a Convolutional Neural Network using Estimators. TensorFlow. https://www.tensorflow.org/tutorials/estimators/cnn (accessed 17 Dec2018).

34 Simonyan K, Vedaldi A, Zisserman A. Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. *arXiv:13126034 [cs]* 2013.http://arxiv.org/abs/1312.6034 (accessed 29 Nov2018).

35 Zeiler MD, Fergus R. Visualizing and Understanding Convolutional Networks. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T (eds). *Computer Vision – ECCV 2014*. Springer International Publishing: Cham, 2014, pp 818–833.

36 Jang M, Horie Y, Shibukawa A, Brake J, Liu Y, Kamali SM *et al.* Wavefront shaping with disorder-engineered metasurfaces. *Nature Photonics* 2018; **12**: 84–90.

37 Kamali SM, Arbabi E, Arbabi A, Faraon A. A review of dielectric optical metasurfaces for wavefront control. *Nanophotonics* 2018; **7**: 1041–1068.

38 Gerke TD, Piestun R. Aperiodic volume optics. *Nature Photonics* 2010; **4**: 188–193.

39 Morizur J-F, Nicholls L, Jian P, Armstrong S, Treps N, Hage B *et al.* Programmable unitary spatial mode manipulation. *Journal of the Optical Society of America A* 2010; **27**: 2524.

40 Oland A, Bansal A, Dannenberg RB, Raj B. Be Careful What You Backpropagate: A Case For Linear Output Activations & Gradient Boosting. *arXiv:170704199 [cs]* 2017.http://arxiv.org/abs/1707.04199 (accessed 8 Jan2019).

41 Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. *ArXiv e-prints* 2014; **1412**: arXiv:1412.6980.

42 Pan SJ, Yang Q. A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering* 2010; **22**: 1345–1359.

43 Horowitz M. 1.1 Computing's energy problem (and what we can do about it). In: *2014 IEEE International Solid-State Circuits Conference Digest of Technical Papers (ISSCC)*. IEEE: San Francisco, CA, USA, 2014, pp 10–14.

44 Rivenson Y, Wang H, Wei Z, de Haan K, Zhang Y, Wu Y *et al.* Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning. *Nat Biomed Eng* 2019; **3**: 466–477.

45 Zhang Y, de Haan K, Rivenson Y, Li J, Delis A, Ozcan A. Digital synthesis of histological stains using micro-structured and multiplexed virtual staining of label-free tissue. *Light: Science & Applications* 2020; **9**. doi:10.1038/s41377-020-0315-y.

46 Chen L-C, Papandreou G, Kokkinos I, Murphy K, Yuille AL. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2018; **40**: 834–848.

47 Rivenson Y, Liu T, Wei Z, Zhang Y, de Haan K, Ozcan A. PhaseStain: the digital staining of label-free quantitative phase microscopy images using deep learning. *Light Sci Appl* 2019; **8**: 23.

48 Amodei D, Ananthanarayanan S, Anubhai R, Bai J, Battenberg E, Case C *et al.* Deep Speech 2 : End-to-End Speech Recognition in English and Mandarin. ; : 10.

49 Zeiler MD, Ranzato M, Monga R, Mao M, Yang K, Le QV *et al.* On rectified linear units for speech processing. In: *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. 2013, pp 3517–3521.

50 Rivenson Y, Wu Y, Ozcan A. Deep learning in holography and coherent imaging. *Light: Science & Applications* 2019; **8**: 1–8.

51 Rivenson Y, Zhang Y, Günaydın H, Teng D, Ozcan A. Phase recovery and holographic image reconstruction using deep learning in neural networks. *Light: Science & Applications* 2018; **7**: 17141–17141.

52 Wu Y, Rivenson Y, Zhang Y, Wei Z, Günaydin H, Lin X *et al.* Extended depth-of-field in holographic imaging using deep-learning-based autofocusing and phase recovery. *Optica* 2018; **5**: 704.

53 Wu Y, Luo Y, Chaudhari G, Rivenson Y, Calis A, de Haan K *et al.* Bright-field holography: cross-modality deep learning enables snapshot 3D imaging with bright-field contrast using a single hologram. *Light: Science & Applications* 2019; **8**. doi:10.1038/s41377-019-0139-9.

54 Li Y, Xue Y, Tian L. Deep speckle correlation: a deep learning approach toward scalable imaging through scattering media. *Optica, OPTICA* 2018; **5**: 1181–1190.

55 LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015; **521**: 436–444.

56 Goodfellow I, Bengio Y, Courville A. *Deep Learning*. MIT Press, 2016.

57 Collobert R, Weston J. A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning. ; : 8.

58 Rahmani B, Loterie D, Konstantinou G, Psaltis D, Moser C. Multimode optical fiber transmission with a deep learning network. *Light Sci Appl* 2018; **7**: 1–11.

59 Bo S, Schmidt F, Eichhorn R, Volpe G. Measurement of anomalous diffusion using recurrent neural networks. *Physical Review E* 2019; **100**. doi:10.1103/PhysRevE.100.010102.

60 Cichos F, Gustavsson K, Mehlig B, Volpe G. Machine learning for active matter. *Nature Machine Intelligence* 2020; **2**: 94–103.

61 de Haan K, Rivenson Y, Wu Y, Ozcan A. Deep-Learning-Based Image Reconstruction and Enhancement in Optical Microscopy. *Proceedings of the IEEE* 2020; **108**: 30–50.

62 Nehme E, Weiss LE, Michaeli T, Shechtman Y. Deep-STORM: super-resolution single-molecule microscopy by deep learning. *Optica* 2018; **5**: 458.

63 Sinha A, Lee J, Li S, Barbastathis G. Lensless computational imaging through deep learning. *Optica* 2017; **4**: 1117.

64 Rivenson Y, Göröcs Z, Günaydin H, Zhang Y, Wang H, Ozcan A. Deep learning microscopy. *Optica* 2017; **4**: 1437.

65 Ouyang W, Aristov A, Lelek M, Hao X, Zimmer C. Deep learning massively accelerates super-resolution localization microscopy. *Nat Biotechnol* 2018; **36**: 460–468.

66 Liu D, Tan Y, Khoram E, Yu Z. Training Deep Neural Networks for the Inverse Design of Nanophotonic Structures. *ACS Photonics* 2018; **5**: 1365–1369.

67 Malkiel I, Mrejen M, Nagler A, Arieli U, Wolf L, Suchowski H. Plasmonic nanostructure design and characterization via Deep Learning. *Light: Science & Applications* 2018; **7**: 60.

68 Ma W, Cheng F, Liu Y. Deep-Learning-Enabled On-Demand Design of Chiral Metamaterials. *ACS Nano* 2018; **12**: 6326–6334.

69 Shen Y, Harris NC, Skirlo S, Prabhu M, Baehr-Jones T, Hochberg M *et al.* Deep learning with coherent nanophotonic circuits. *Nature Photon* 2017; **11**: 441–446.

70 Feldmann J, Youngblood N, Wright CD, Bhaskaran H, Pernice WHP. All-optical spiking neurosynaptic networks with self-learning capabilities. *Nature* 2019; **569**: 208–214.

71 Mehrabian A, Al-Kabani Y, Sorger VJ, El-Ghazawi T. PCNNA: A Photonic Convolutional Neural Network Accelerator. In: *2018 31st IEEE International System-on-Chip Conference (SOCC)*. 2018, pp 169–173.

72 Miscuglio M, Mehrabian A, Hu Z, Azzam SI, George J, Kildishev AV *et al.* All-optical nonlinear activation function for photonic neural networks [Invited]. *Optical Materials Express* 2018; **8**: 3851.

73 Psaltis D, Brady D, Gu X-G, Lin S. Holography in artificial neural networks. *Nature* 1990; **343**: 325–330.

74 Shastri BJ, Tait AN, Ferreira de Lima T, Nahmias MA, Peng H-T, Prucnal PR. Neuromorphic Photonics, Principles of. In: Meyers RA (ed). *Encyclopedia of Complexity and Systems Science*. Springer Berlin Heidelberg: Berlin, Heidelberg, 2018, pp 1–37.

75 Estakhri NM, Edwards B, Engheta N. Inverse-designed metastructures that solve equations. *Science* 2019; **363**: 1333–1338.

76 Chang J, Sitzmann V, Dun X, Heidrich W, Wetzstein G. Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification. *Scientific Reports* 2018; **8**: 12324.

77 Lin X, Rivenson Y, Yardimci NT, Veli M, Luo Y, Jarrahi M *et al.* All-optical machine learning using diffractive deep neural networks. *Science* 2018; **361**: 1004–1008.

78 Mengu D, Luo Y, Rivenson Y, Ozcan A. Analysis of Diffractive Optical Neural Networks and Their Integration With Electronic Neural Networks. *IEEE J Select Topics Quantum Electron* 2020; **26**: 1–14.

79 Li J, Mengu D, Luo Y, Rivenson Y, Ozcan A. Class-specific differential detection in diffractive optical neural networks improves inference accuracy. *AP* 2019; **1**: 046001.

80 Luo Y, Mengu D, Yardimci NT, Rivenson Y, Veli M, Jarrahi M *et al.* Design of task-specific optical systems using broadband diffractive neural networks. *Light Sci Appl* 2019; **8**: 112.

81 Sande GV der, Brunner D, Soriano MC. Advances in photonic reservoir computing. *Nanophotonics* 2017; **6**: 561–576.

82 Marandi A, Wang Z, Takata K, Byer RL, Yamamoto Y. Network of time-multiplexed optical parametric oscillators as a coherent Ising machine. *Nature Photonics* 2014; **8**: 937–942.

83 Penkovsky B, Porte X, Jacquot M, Larger L, Brunner D. Coupled Nonlinear Delay Systems as Deep Convolutional Neural Networks. *Physical Review Letters* 2019; **123**. doi:10.1103/PhysRevLett.123.054101.

84 Bueno J, Maktoobi S, Froehly L, Fischer I, Jacquot M, Larger L *et al.* Reinforcement learning in a large-scale photonic recurrent neural network. *Optica* 2018; **5**: 756.

85 George JK, Mehrabian A, Amin R, Meng J, de Lima TF, Tait AN *et al.* Neuromorphic photonics with electro-absorption modulators. *Optics Express* 2019; **27**: 5181.

86 Ribeiro A, Ruocco A, Vanacker L, Bogaerts W. Demonstration of a 4 × 4-port universal linear circuit. *Optica* 2016; **3**: 1348.

87 LeCun Y, Bottou L, Bengio Y, Ha P. Gradient-Based Learning Applied to Document Recognition. 1998; : 46.

88 Xiao H, Rasul K, Vollgraf R. Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms. *arXiv:170807747 [cs, stat]* 2017.http://arxiv.org/abs/1708.07747 (accessed 13 May2020).

89 Delen N, Hooker B. Free-space beam propagation between arbitrarily oriented planes based on full diffraction theory: a fast Fourier transform approach. *Journal of the Optical Society of America A* 1998; **15**: 857.

90 Matsushima K, Schimmel H, Wyrowski F. Fast calculation method for optical diffraction on tilted planes by use of the angular spectrum of plane waves. *Journal of the Optical Society of America A* 2003; **20**: 1755.

91 Marcucci G, Pierangeli D, Conti C. Theory of neuromorphic computing by waves: machine learning by rogue waves, dispersive shocks, and solitons. *arXiv:191207044 [cond-mat, physics:nlin, physics:physics]* 2019.http://arxiv.org/abs/1912.07044 (accessed 12 Jun2020).

92 Hughes TW, Williamson IAD, Minkov M, Fan S. Wave physics as an analog recurrent neural network. *Science Advances* 2019; **5**: eaay6946.

93 Wu Y, Rivenson Y, Wang H, Luo Y, Ben-David E, Bentolila LA *et al.* Three-dimensional virtual refocusing of fluorescence microscopy images using deep learning. *Nature Methods* 2019; **16**: 1323–1331.

94 Malkiel I, Mrejen M, Nagler A, Arieli U, Wolf L, Suchowski H. Plasmonic nanostructure design and characterization via Deep Learning. *Light Sci Appl* 2018; **7**: 60.

95 Miscuglio M, Mehrabian A, Hu Z, Azzam SI, George J, Kildishev AV *et al.* All-optical nonlinear activation function for photonic neural networks [Invited]. *Optical Materials Express* 2018; **8**: 3851.

96 Hughes TW, Minkov M, Shi Y, Fan S. Training of photonic neural networks through in situ backpropagation and gradient measurement. *Optica* 2018; **5**: 864.

97 Prabhu M, Roques-Carmes C, Shen Y, Harris N, Jing L, Carolan J *et al.* A Recurrent Ising Machine in a Photonic Integrated Circuit. *arXiv:190913877 [physics]* 2019.http://arxiv.org/abs/1909.13877 (accessed 30 May2020).

98 Rahman SS, Li J, Mengu D, Rivenson Y, Ozcan A. Ensemble learning of diffractive optical networks. *arXiv:200906869 [cs, eees, physics]*; : 22.

99 Li J, Mengu D, Yardimci NT, Luo Y, Li X, Veli M *et al.* Machine Vision using Diffractive Spectral Encoding. *arXiv:200511387 [cs, eess, physics]* 2020.

100 Kulce O, Mengu D, Rivenson Y, Ozcan A. All-Optical Information Processing Capacity of Diffractive Surfaces. *arXiv:200712813 [eees, cs, cs, physics]*; : 28.

101 Veli M, Mengu D, Yardimci NT, Luo Y, Li J, Rivenson Y *et al.* Terahertz Pulse Shaping Using Diffractive Legos. *arXiv:200616599 [cs, physics]*.

102 Wagner K, Psaltis D. Multilayer optical learning networks. *Appl Opt, AO* 1987; **26**: 5061–5076.

103 Mengu D, Zhao Y, Yardimci NT, Rivenson Y, Jarrahi M, Ozcan A. Misalignment resilient diffractive optical networks. *Nanophotonics* 2020; **9**: 4207–4219.

104 Guo C, Pleiss G, Sun Y, Weinberger KQ. On Calibration of Modern Neural Networks. *JMLR.org* 2017; **70**.

105 Laha A, Chemmengath SA, Agrawal P, Khapra M, Sankaranarayanan K, Ramaswamy HG. On Controllable Sparse Alternatives to Softmax. In: Bengio S, Wallach H, Larochelle H, Grauman K, Cesa-Bianchi N, Garnett R (eds). *Advances in Neural Information Processing Systems 31*. Curran Associates, Inc., 2018, pp 6422–6432.

106 Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. *arXiv:14126980 [cs]* 2014.http://arxiv.org/abs/1412.6980 (accessed 7 Jun2019).

107 Pendry JB. Negative Refraction Makes a Perfect Lens. *Physical Review Letters* 2000; **85**: 3966–3969.

108 Cubukcu E, Aydin K, Ozbay E, Foteinopoulo S, Soukoulis CM. Negative refraction by photonic crystals. *Nature* 2003; **423**: 604–605.

109 Fang N. Sub-Diffraction-Limited Optical Imaging with a Silver Superlens. *Science* 2005; **308**: 534–537.

110 Jacob Z, Alekseyev LV, Narimanov E. Optical Hyperlens: Far-field imaging beyond the diffraction limit. 2006; : 10.

111 Engheta N. Circuits with Light at Nanoscales: Optical Nanocircuits Inspired by Metamaterials. *Science* 2017; **317**: 1698–1702.

112 Liu Z, Lee H, Xiong Y, Sun C, Zhang X. Far-Field Optical Hyperlens Magnifying Sub-Diffraction-Limited Objects. *Science* 2007; **315**: 1686–1686.

113 MacDonald KF, Sámson ZL, Stockman MI, Zheludev NI. Ultrafast active plasmonics. *Nature Photonics* 2009; **3**: 55–58.

114 Lin D, Fan P, Hasman E, Brongersma ML. Dielectric gradient metasurface optical elements. *Science* 2014; **345**: 298–302.

115 Yu N, Capasso F. Flat optics with designer metasurfaces. *Nature Materials* 2014; **13**: 139–150.

116 Kuznetsov AI, Miroshnichenko AE, Brongersma ML, Kivshar YS, Luk'yanchuk B. Optically resonant dielectric nanostructures. *Science* 2016; **354**: aag2472.

117 Shalaev VM. Optical negative-index metamaterials. *Nature Photonics* 2007; **1**: 41–48.

118 Chen H-T, Taylor AJ, Yu N. A review of metasurfaces: physics and applications. *Reports on Progress in Physics* 2016; **79**: 076401.

119 Smith DR. Metamaterials and Negative Refractive Index. *Science* 2004; **305**: 788–792.

120 Yu N, Genevet P, Aieta F, Kats MA, Blanchard R, Aoust G *et al.* Flat Optics: Controlling Wavefronts With Optical Antenna Metasurfaces. *IEEE Journal of Selected Topics in Quantum Electronics* 2013; **19**: 4700423–4700423.

121 Maier SA, Kik PG, Atwater HA, Meltzer S, Harel E, Koel BE *et al.* Local detection of electromagnetic energy transport below the diffraction limit in metal nanoparticle plasmon waveguides. *Nature Materials* 2003; **2**: 229–232.

122 Alù A, Engheta N. Achieving transparency with plasmonic and metamaterial coatings. *Physical Review E* 2005; **72**. doi:10.1103/PhysRevE.72.016623.

123 Schurig D, Mock JJ, Justice BJ, Cummer SA, Pendry JB, Starr AF *et al.* Metamaterial Electromagnetic Cloak at Microwave Frequencies. *Science* 2006; **314**: 977–980.

124 Pendry JB. Controlling Electromagnetic Fields. *Science* 2006; **312**: 1780–1782.

125 Cai W, Chettiar UK, Kildishev AV, Shalaev VM. Optical cloaking with metamaterials. *Nature Photonics* 2007; **1**: 224–227.

126 Valentine J, Li J, Zentgraf T, Bartal G, Zhang X. An optical cloak made of dielectrics. *Nature Materials* 2009; **8**: 568–571.

127 Narimanov EE, Kildishev AV. Optical black hole: Broadband omnidirectional light absorber. *Applied Physics Letters* 2009; **95**: 041106.

128 Oulton RF, Sorger VJ, Zentgraf T, Ma R-M, Gladden C, Dai L *et al.* Plasmon lasers at deep subwavelength scale. *Nature* 2009; **461**: 629–632.

129 Zhao Y, Belkin MA, Alù A. Twisted optical metamaterials for planarized ultrathin broadband circular polarizers. *Nature Communications* 2012; **3**. doi:10.1038/ncomms1877.

130 Watts CM, Shrekenhamer D, Montoya J, Lipworth G, Hunt J, Sleasman T *et al.* Terahertz compressive imaging with metamaterial spatial light modulators. *Nature Photonics* 2014; **8**: 605–609.

131 Qian C, Lin X, Lin X, Xu J, Sun Y, Li E *et al.* Performing optical logic operations by a diffractive neural network. *Light: Science & Applications* 2020; **9**. doi:10.1038/s41377-020-0303-2.

132 Mengu D, Zhao Y, Yardimci NT, Rivenson Y, Jarrahi M, Ozcan A. Misalignment resilient diffractive optical networks. *Nanophotonics* 2020; **0**. doi:10.1515/nanoph-2020-0291.

133 Esmer GB, Uzunov V, Onural L, Ozaktas HM, Gotchev A. Diffraction field computation from arbitrarily distributed data points in space. *Signal Processing: Image Communication* 2007; **22**: 178–187.

134 Goodman JW. *Introduction to Fourier Optics*. Roberts and Company Publishers, 2005.

135 Zhang Z, You Z, Chu D. Fundamentals of phase-only liquid crystal on silicon (LCOS) devices. *Light: Science & Applications* 2014; **3**: e213–e213.

136 Moon TK, Sterling WC. Mathematical Methods and Algorithms for Signal Processing. 2000.

137 CIFAR-10 and CIFAR-100 datasets. https://www.cs.toronto.edu/~kriz/cifar.html (accessed 15 Jul2020).

138 Wetzstein G, Ozcan A, Gigan S, Fan S, Englund D, Soljačić M *et al.* Inference in artificial intelligence with deep optics and photonics. *Nature* 2020; **588**: 39–47.

139 Ozaktas HM, Mendlovic D, Kutay MA, Zalevsky Z. *The Fractional Fourier Transform: With Applications in Optics and Signal Processing*. Wiley, 2001.

140 Goodman JW. *Introduction to Fourier Optics*. Roberts and Company Publishers, 2005.

141 Athale R, Psaltis D. Optical Computing: Past and Future. *Optics & Photonics News* 2016; **27**: 32.

142 Solli DR, Jalali B. Analog optical computing. *Nature Photon* 2015; **9**: 704–706.

143 Zangeneh-Nejad F, Sounas DL, Alù A, Fleury R. Analogue computing with metamaterials. *Nature Reviews Materials* 2021; **6**: 207–225.

144 Miller DAB. Self-configuring universal linear optical component [Invited]. *Photon Res* 2013; **1**: 1.

145 Reck M, Zeilinger A, Bernstein HJ, Bertani P. Experimental realization of any discrete unitary operator. *Phys Rev Lett* 1994; **73**: 58–61.

146 Goodman JW, Dias AR, Woody LM. Fully parallel, high-speed incoherent optical method for performing discrete Fourier transforms. *Opt Lett* 1978; **2**: 1.

147 Slavík R, Park Y, Ayotte N, Doucet S, Ahn T-J, LaRochelle S *et al.* Photonic temporal integrator for all-optical computing. *Opt Express* 2008; **16**: 18202.

148 Goodman JW, Woody LM. Method for performing complex-valued linear operations on complex-valued data using incoherent light. *Appl Opt, AO* 1977; **16**: 2611–2612.

149 Farhat NH, Psaltis D, Prata A, Paek E. Optical implementation of the Hopfield model. *Appl Opt, AO* 1985; **24**: 1469–1475.

150 Athale RA, Collins WC. Optical matrix–matrix multiplier based on outer product decomposition. *Appl Opt* 1982; **21**: 2089.

151 Sawchuk AA, Strand TC. Digital optical computing. *Proceedings of the IEEE* 1984; **72**: 758–779.

152 Moeini MM, Sounas DL. Discrete space optical signal processing. *Optica* 2020; **7**: 1325.

153 Silva A, Monticone F, Castaldi G, Galdi V, Alù A, Engheta N. Performing Mathematical Operations with Metamaterials. *Science* 2014; **343**: 160–163.

154 Miscuglio M, Hu Z, Li S, George JK, Capanna R, Dalir H *et al.* Massively parallel amplitude-only Fourier neural network. *Optica* 2020; **7**: 1812.

155 Yu FTS, Jutamulia S. *Optical Signal Processing, Computing, and Neural Networks*. Wiley, 1992.

156 Duport F, Schneider B, Smerieri A, Haelterman M, Massar S. All-optical reservoir computing. *Opt Express, OE* 2012; **20**: 22783–22795.

157 Knill E, La R. A scheme for ef®cient quantum computation with linear optics. 2001; **409**: 7.

158 Shen Y, Harris NC, Skirlo S, Prabhu M, Baehr-Jones T, Hochberg M *et al.* Deep learning with coherent nanophotonic circuits. *Nature Photonics* 2017; **11**: 441–446.

159 Pendry JB. Negative Refraction Makes a Perfect Lens. *Phys Rev Lett* 2000; **85**: 3966–3969.

160 Pendry JB, Schurig D, Smith DR. Controlling Electromagnetic Fields. *Science* 2006; **312**: 1780–1782.

161 Cai W, Chettiar UK, Kildishev AV, Shalaev VM. Optical cloaking with metamaterials. *Nature Photonics* 2007; **1**: 224–227.

162 Valentine J, Li J, Zentgraf T, Bartal G, Zhang X. An optical cloak made of dielectrics. *Nature Materials* 2009; **8**: 568–571.

163 Oulton RF, Sorger VJ, Zentgraf T, Ma R-M, Gladden C, Dai L *et al.* Plasmon lasers at deep subwavelength scale. *Nature* 2009; **461**: 629–632.

164 Estakhri NM, Edwards B, Engheta N. Inverse-designed metastructures that solve equations. *Science* 2019; **363**: 1333–1338.

165 Lin X, Rivenson Y, Yardimci NT, Veli M, Luo Y, Jarrahi M *et al.* All-optical machine learning using diffractive deep neural networks. *Science* 2018; **361**: 1004–1008.

166 Li J, Mengu D, Yardimci NT, Luo Y, Li X, Veli M *et al.* Spectrally encoded single-pixel machine vision using diffractive networks. *Science Advances* 2021; **7**: eabd7690.

167 Rahman MSS, Li J, Mengu D, Rivenson Y, Ozcan A. Ensemble learning of diffractive optical networks. *Light: Science & Applications* 2021; **10**: 14.

168 Mengu D, Zhao Y, Yardimci NT, Rivenson Y, Jarrahi M, Ozcan A. Misalignment resilient diffractive optical networks. *Nanophotonics* 2020; **1**. doi:10.1515/nanoph-2020-0291.

169 Veli M, Mengu D, Yardimci NT, Luo Y, Li J, Rivenson Y *et al.* Terahertz Pulse Shaping Using Diffractive Surfaces. *arXiv:200616599 [physics]* 2020.http://arxiv.org/abs/2006.16599 (accessed 1 Jan2021).

170 Mengu D, Rivenson Y, Ozcan A. Scale-, Shift-, and Rotation-Invariant Diffractive Optical Networks. *ACS Photonics* 2021; **8**: 324–334.

171 Luo Y, Mengu D, Yardimci NT, Rivenson Y, Veli M, Jarrahi M *et al.* Design of task-specific optical systems using broadband diffractive neural networks. *Light: Science & Applications* 2019; **8**: 112.

172 Kulce O, Mengu D, Rivenson Y, Ozcan A. All-optical information-processing capacity of diffractive surfaces. *Light: Science & Applications* 2021; **10**: 25.

173 Zuo Y, Li B, Zhao Y, Jiang Y, Chen Y-C, Chen P *et al.* All-optical neural network with nonlinear activation functions. *Optica, OPTICA* 2019; **6**: 1132–1137.

174 Dinc NU, Link to external site this link will open in a new window, Lim J, Kakkava E, Moser C, Psaltis D. Computer generated optical volume elements by additive manufacturing. *Nanophotonics* 2020; **9**: 4173–4181.

175 Kulce O, Onural L, Ozaktas HM. Evaluation of the validity of the scalar approximation in optical wave propagation using a systems approach and an accurate digital electromagnetic model. *Journal of Modern Optics* 2016; **63**: 2382–2391.

176 Kulce O, Onural L. Power Spectrum Equalized Scalar Representation of Wide-Angle Optical Field Propagation. *J Math Imaging Vis* 2018; **60**: 1246–1260.

177 Kulce O, Onural L. Generation of a polarized optical field from a given scalar field for wide-viewing-angle holographic displays. *Optics and Lasers in Engineering* 2021; **137**: 106344.

178 Moon TK, Stirling WC. *Mathematical Methods and Algorithms for Signal Processing*. Prentice Hall, 2000.

179 Oppenheim AV. *Discrete-time signal processing*. Pearson Education India, 1999.

180 Hayes MH. *Statistical Digital Signal Processing and Modeling*. 1st edition. Wiley: New York, 1996.

181 Ishikawa N, Sugiura S, Hanzo L. 50 Years of Permutation, Spatial and Index Modulation: From Classic RF to Visible Light Communications and Data Storage. *IEEE Commun Surv Tutorials* 2018; **20**: 1905–1938.

182 Ishimura S, Kikuchi K. Multi-dimensional permutation-modulation format for coherent optical communications. *Opt Express* 2015; **23**: 15587.

183 Huang H, He X, Xiang Y, Wen W, Zhang Y. A compression-diffusion-permutation strategy for securing image. *Signal Processing* 2018; **150**: 183–190.

184 Anxiao Jiang, Mateescu R, Schwartz M, Bruck J. Rank Modulation for Flash Memories. *IEEE Trans Inform Theory* 2009; **55**: 2659–2673.

185 Huang X, Ye G, Chai H, Xie O. Compression and encryption for remote sensing image using chaotic system. *Security and Communication Networks* 2015; **8**: 3659–3666.

186 Carolan J, Harrold C, Sparrow C, Martín-López E, Russell NJ, Silverstone JW *et al.* Universal linear optics. *Science* 2015; **349**: 711–716.

187 Spanke RA, Benes VE. N-stage planar optical permutation network. *Appl Opt* 1987; **26**: 1226.

188 Djavid M, Dastjerdi MHT, Philip MR, Choudhary DD, Pham TT, Khreishah A *et al.* Photonic crystal-based permutation switch for optical networks. *Photon Netw Commun* 2018; **35**: 90–96.

189 Kobolla H, Sauer F, Volkel R. Holographic Tandem Arrays. In: Morris GM (ed). . Paris, France, 1989, p 146.

190 Robertson B, Restall EJ, Taghizadeh MR, Walker AC. Space-variant holographic optical elements in dichromated gelatin. *Appl Opt* 1991; **30**: 2368.

191 Kobolla H, Sheridan JT, Gluch E, Schmidt J, Völkel R, Schwider J *et al.* Holographic 2D Mixed Polarization Deflection Elements. *Journal of Modern Optics* 1993; **40**: 613–624.

192 Hutley MC, Savander P, Schrader M. The use of microlenses for making spatially variant optical interconnections. *Pure Appl Opt* 1992; **1**: 337–346.

193 Jahns J, Däschner W. Optical cyclic shifter using diffractive lenslet arrays. *Optics Communications* 1990; **79**: 407–410.

194 Sauer F, Jahns J, Nijander CR, Feldblum AY, Townsend WP. Refractive-diffractive micro-optics for permutation interconnects. *Opt Eng* 1994; **33**: 1550.

195 Tarable A, Malandrino F, Dossi L, Nebuloni R, Virone G, Nordio A. Meta-Surface Optimization in 6G Sub-THz Communications. In: *2020 IEEE International Conference on Communications Workshops (ICC Workshops)*. 2020, pp 1–6.

196 Xu X, Chen Q, Mu X, Liu Y, Jiang H. Graph-Embedded Multi-Agent Learning for Smart Reconfigurable THz MIMO-NOMA Networks. *IEEE Journal on Selected Areas in Communications* 2022; **40**: 259–275.

197 King BM, Neifeld MA. Sparse modulation coding for increased capacity in volume holographic storage. *Appl Opt* 2000; **39**: 6681.

198 Enayatifar R, Abdullah AH, Isnin IF, Altameem A, Lee M. Image encryption using a synchronous permutation-diffusion technique. *Optics and Lasers in Engineering* 2017; **90**: 146–154.

199 Patidar V, Pareek NK, Purohit G, Sud KK. A robust and secure chaotic standard map based pseudorandom permutation-substitution scheme for image encryption. *Optics Communications* 2011; **284**: 4331–4339.

200 Bai B, Luo Y, Gan T, Hu J, Li Y, Zhao Y *et al.* To image, or not to image: Class-specific diffractive cameras with all-optical erasure of undesired objects. 2022. doi:10.48550/ARXIV.2205.13122.

201 Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Communications of the ACM* 2017; **60**: 84–90.

202 LeCun Y, Bottou L, Bengio Y, Ha P. Gradient-Based Learning Applied to Document Recognition. *Proceedings of the IEEE* 1998; **86**: 2278–2374.

203 Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF (eds). *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Springer International Publishing: Cham, 2015, pp 234–241.

204 Collobert R, Weston J. A unified architecture for natural language processing: deep neural networks with multitask learning. In: *Proceedings of the 25th international conference on Machine learning - ICML '08*. ACM Press: Helsinki, Finland, 2008, pp 160–167.

205 Goldberg Y. Neural Network Methods for Natural Language Processing. *Synthesis Lectures on Human Language Technologies* 2017; **10**: 1–309.

206 Goodfellow I, Bengio Y, Courville A. *Deep Learning*. The MIT Press, 2016.

207 Rivenson Y, Ceylan Koydemir H, Wang H, Wei Z, Ren Z, Günaydın H *et al.* Deep Learning Enhanced Mobile-Phone Microscopy. *ACS Photonics* 2018; **5**: 2354–2364.

208 de Haan K, Ballard ZS, Rivenson Y, Wu Y, Ozcan A. Resolution enhancement in scanning electron microscopy using deep learning. *Sci Rep* 2019; **9**: 12050.

209 Nehme E, Freedman D, Gordon R, Ferdman B, Weiss LE, Alalouf O *et al.* DeepSTORM3D: dense 3D localization microscopy and PSF design by deep learning. *Nature Methods* 2020; **17**: 734–740.

210 Nguyen T, Nguyen T, Xue Y, Li Y, Tian L, Tian L *et al.* Deep learning approach for Fourier ptychography microscopy. *Opt Express, OE* 2018; **26**: 26470–26484.

211 Jo Y, Cho H, Lee SY, Choi G, Kim G, Min H *et al.* Quantitative Phase Imaging and Artificial Intelligence: A Review. *IEEE Journal of Selected Topics in Quantum Electronics* 2019; **25**: 1–14.

212 Park Y, Depeursinge C, Popescu G. Quantitative phase imaging in biomedicine. *Nature Photon* 2018; **12**: 578–589.

213 Goswami N, He YR, Deng Y-H, Oh C, Sobh N, Valera E *et al.* Label-free SARS-CoV-2 detection and classification using phase imaging with computational specificity. *Light Sci Appl* 2021; **10**: 176.

214 Borhani N, Kakkava E, Moser C, Psaltis D. Learning to see through multimode fibers. *Optica* 2018; **5**: 960.

215 Bianco V, Mazzeo PL, Paturzo M, Distante C, Ferraro P. Deep learning assisted portable IR active imaging sensor spots and identifies live humans through fire. *Optics and Lasers in Engineering* 2020; **124**: 105818.

216 You S, Chaney EJ, Tu H, Sun Y, Sinha S, Boppart SA. Label-Free Deep Profiling of the Tumor Microenvironment. *Cancer Res* 2021; **81**: 2534–2544.

217 Yoon J, Jo Y, Kim M, Kim K, Lee S, Kang S-J *et al.* Identification of non-activated lymphocytes using three-dimensional refractive index tomography and machine learning. *Sci Rep* 2017; **7**: 6654.

218 Li J, Garfinkel J, Zhang X, Wu D, Zhang Y, de Haan K *et al.* Biopsy-free in vivo virtual histology of skin using deep learning. *Light Sci Appl* 2021; **10**: 233.

219 Wu J, Cao L, Cao L, Barbastathis G, Barbastathis G. DNN-FZA camera: a deep learning approach toward broadband FZA lensless imaging. *Opt Lett, OL* 2021; **46**: 130–133.

220 Ma W, Cheng F, Xu Y, Wen Q, Liu Y. Probabilistic Representation and Inverse Design of Metamaterials Based on a Deep Generative Model with Semi-Supervised Learning Strategy. *Advanced Materials* 2019; **31**: 1901111.

221 Schwartz E, Giryes R, Bronstein AM. DeepISP: Towards Learning an End-to-End Image Processing Pipeline. *IEEE Trans on Image Process* 2019; **28**: 912–923.

222 Sitzmann V, Diamond S, Peng Y, Dun X, Boyd S, Heidrich W *et al.* End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging. *ACM Transactions on Graphics* 2018; **37**: 1–13.

223 Ballard Z, Brown C, Madni AM, Ozcan A. Machine learning and computation-enabled intelligent sensor design. *Nat Mach Intell* 2021; **3**: 556–565.

224 Joung H-A, Ballard ZS, Wu J, Tseng DK, Teshome H, Zhang L *et al.* Point-of-Care Serodiagnostic Test for Early-Stage Lyme Disease Using a Multiplexed Paper-Based Immunoassay and Machine Learning. *ACS Nano* 2020; **14**: 229–240.

225 Metzler CA, Ikoma H, Peng Y, Wetzstein G. Deep Optics for Single-Shot High-Dynamic-Range Imaging. In: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE: Seattle, WA, USA, 2020, pp 1372–1382.

226 Hu L, Hu S, Gong W, Si K, Gong W, Si K *et al.* Learning-based Shack-Hartmann wavefront sensor for high-order aberration detection. *Opt Express, OE* 2019; **27**: 33504–33517.

227 Schmeisser M, Heisen BC, Luettich M, Busche B, Hauer F, Koske T *et al.* Parallel, distributed and GPU computing technologies in single-particle electron microscopy. *Acta Crystallogr D Biol Crystallogr* 2009; **65**: 659–671.

228 Xu X, Tan M, Corcoran B, Wu J, Boes A, Nguyen TG *et al.* 11 TOPS photonic convolutional accelerator for optical neural networks. *Nature* 2021; **589**: 44–51.

229 Tait AN, Nahmias MA, Shastri BJ, Prucnal PR. Broadcast and Weight: An Integrated Network For Scalable Photonic Spike Processing. *Journal of Lightwave Technology* 2014; **32**: 4029–4041.

230 Tait AN, Ferreira de Lima T, Nahmias MA, Miller HB, Peng H-T, Shastri BJ *et al.* Silicon Photonic Modulator Neuron. *Phys Rev Applied* 2019; **11**: 064043.

231 Goi E, Chen X, Zhang Q, Cumming BP, Schoenhardt S, Luan H *et al.* Nanoprinted high-neuron-density optical linear perceptrons performing near-infrared inference on a CMOS chip. *Light Sci Appl* 2021; **10**: 40.

232 Bai B, Luo Y, Gan T, Hu J, Li Y, Zhao Y *et al.* To image, or not to image: Class-specific diffractive cameras with all-optical erasure of undesired objects. ; : 31.

233 Kulce O, Mengu D, Rivenson Y, Ozcan A. All-optical synthesis of an arbitrary linear transformation using diffractive surfaces. *Light Sci Appl* 2021; **10**: 196.

234 Isil C, Mengu D, Zhao Y, Tabassum A, Li J, Luo Y *et al.* Super-resolution image display using diffractive decoders. 2022. doi:10.48550/ARXIV.2206.07281.

235 Mengu D, Ozcan A. All-Optical Phase Recovery: Diffractive Computing for Quantitative Phase Imaging. *Advanced Optical Materials*; **n/a**: 2200281.

236 Luo Y, Zhao Y, Li J, Çetintaş E, Rivenson Y, Jarrahi M *et al.* Computational imaging without a computer: seeing through random diffusers at the speed of light. *eLight* 2022; **2**: 4.

237 Tuchin VV, others. Tissue optics. Society of Photo-Optical Instrumentation Engineers (SPIE), 2015.

238 Diaspro A. *Optical fluorescence microscopy: From the spectral to the nano dimension*. Springer Science & Business Media, 2010.

239 Popescu G. *Quantitative phase imaging of cells and tissues*. McGraw-Hill: New York, 2011.

240 Zernike F. How I Discovered Phase Contrast. *Science* 1955; **121**: 345–349.

241 Lang W. *Nomarski differential interference-contrast microscopy*. Carl Zeiss, 1982.

242 Shaked NT, Rinehart MT, Wax A. Dual-interference-channel quantitative-phase microscopy of live cell dynamics. *Opt Lett* 2009; **34**: 767.

243 Memmolo P, Iannone M, Ventre M, Netti PA, Finizio A, Paturzo M *et al.* On the holographic 3D tracking of in vitro cells characterized by a highly-morphological change. *Opt Express* 2012; **20**: 28485.

244 Park H, Lee S, Ji M, Kim K, Son Y, Jang S *et al.* Measuring cell surface area and deformability of individual human red blood cells over blood storage using quantitative phase imaging. *Sci Rep* 2016; **6**: 34257.

245 Popescu G, Deflores LP, Vaughan JC, Badizadegan K, Iwai H, Dasari RR *et al.* Fourier phase microscopy for investigation of biological structures and dynamics. *Opt Lett* 2004; **29**: 2503.

246 Ikeda T, Popescu G, Dasari RR, Feld MS. Hilbert phase microscopy for investigating fast dynamics in transparent systems. *Opt Lett* 2005; **30**: 1165.

247 Marquet P, Rappaz B, Magistretti PJ, Cuche E, Emery Y, Colomb T *et al.* Digital holographic microscopy: a noninvasive contrast imaging technique allowing quantitative visualization of living cells with subwavelength axial accuracy. *Opt Lett* 2005; **30**: 468.

248 Greenbaum A, Zhang Y, Feizi A, Chung P-L, Luo W, Kandukuri SR *et al.* Wide-field computational imaging of pathology slides using lens-free on-chip microscopy. *Sci Transl Med* 2014; **6**. doi:10.1126/scitranslmed.3009850.

249 Cuche E, Bevilacqua F, Depeursinge C. Digital holography for quantitative phase-contrast imaging. *Opt Lett* 1999; **24**: 291.

250 Greenbaum A, Luo W, Su T-W, Göröcs Z, Xue L, Isikman SO *et al.* Imaging without lenses: achievements and remaining challenges of wide-field on-chip microscopy. *Nat Methods* 2012; **9**: 889–895.

251 Matlock A, Tian L. High-throughput, volumetric quantitative phase imaging with multiplexed intensity diffraction tomography. *Biomed Opt Express* 2019; **10**: 6432.

252 Doblas A, Sánchez-Ortiga E, Martínez-Corral M, Saavedra G, Andrés P, Garcia-Sucerquia J. Shift-variant digital holographic microscopy: inaccuracies in quantitative phase imaging. *Opt Lett* 2013; **38**: 1352.

253 Bon P, Maucort G, Wattellier B, Monneret S. Quadriwave lateral shearing interferometry for quantitative phase microscopy of living cells. *Opt Express* 2009; **17**: 13080.

254 Wang Z, Millet L, Mir M, Ding H, Unarunotai S, Rogers J *et al.* Spatial light interference microscopy (SLIM). 2011; : 11.

255 Nguyen TH, Kandel ME, Rubessa M, Wheeler MB, Popescu G. Gradient light interference microscopy for 3D imaging of unlabeled specimens. *Nat Commun* 2017; **8**: 210.

256 Majeed H, Ma L, Lee YJ, Kandel M, Min E, Jung W *et al.* Magnified Image Spatial Spectrum (MISS) microscopy for nanometer and millisecond scale label-free imaging. *Opt Express* 2018; **26**: 5423.

257 Vijayakumar A, Kashter Y, Kelner R, Rosen J. Coded aperture correlation holography–a new type of incoherent digital holograms. *Opt Express* 2016; **24**: 12430.

258 Tian L, Waller L. 3D intensity and phase imaging from light field measurements in an LED array microscope. *Optica* 2015; **2**: 104.

259 Tian L, Li X, Ramchandran K, Waller L. Multiplexed coded illumination for Fourier Ptychography with an LED array microscope. *Biomed Opt Express, BOE* 2014; **5**: 2376–2389.

260 Lim J, Ayoub AB, Antoine EE, Psaltis D. High-fidelity optical diffraction tomography of multiple scattering samples. *Light: Science & Applications* 2019; **8**. doi:10.1038/s41377-019-0195-1.

261 Kelner R, Rosen J. Spatially incoherent single channel digital Fourier holography. *Opt Lett* 2012; **37**: 3723.

262 Lee K, Park Y. Exploiting the speckle-correlation scattering matrix for a compact reference-free holographic image sensor. *Nat Commun* 2016; **7**: 13359.

263 Cotte Y, Toy F, Jourdain P, Pavillon N, Boss D, Magistretti P *et al.* Marker-free phase nanoscopy. *Nature Photon* 2013; **7**: 113–117.

264 Popescu G, Badizadegan K, Dasari RR, Feld MS. Observation of dynamic subdomains in red blood cells. *J Biomed Opt* 2006; **11**: 040503.

265 Popescu G, Park Y, Lue N, Best-Popescu C, Deflores L, Dasari RR *et al.* Optical imaging of cell mass and growth dynamics. *American Journal of Physiology-Cell Physiology* 2008; **295**: C538–C544.

266 Mitchell S, Roy K, Zangle TA, Hoffmann A. Nongenetic origins of cell-to-cell variability in B lymphocyte proliferation. *Proc Natl Acad Sci USA* 2018; **115**: E2888–E2897.

267 Uttam S, Pham HV, LaFace J, Leibowitz B, Yu J, Brand RE *et al.* Early Prediction of Cancer Progression by Depth-Resolved Nanoscale Mapping of Nuclear Architecture from Unstained Tissue Specimens. *Cancer Res* 2015; **75**: 4718–4727.

268 Roitshtain D, Wolbromsky L, Bal E, Greenspan H, Satterwhite LL, Shaked NT. Quantitative phase microscopy spatial signatures of cancer cells. *Cytometry* 2017; **91**: 482–493.

269 Watanabe E, Hoshiba T, Javidi B. High-precision microscopic phase imaging without phase unwrapping for cancer cell identification. *Opt Lett* 2013; **38**: 1319.

270 Rubin M, Stein O, Turko NA, Nygate Y, Roitshtain D, Karako L *et al.* TOP-GAN: Stain-free cancer cell classification using deep learning with a small training set. *Medical Image Analysis* 2019; **57**: 176–185.

271 Shaked NT, Zhu Y, Badie N, Bursac N, Wax A. Reflective interferometric chamber for quantitative phase imaging of biological sample dynamics. *J Biomed Opt* 2010; **15**: 030503.

272 Shaked NT, Zhu Y, Rinehart MT, Wax A. Two-step-only phase-shifting interferometry with optimized detector bandwidth for microscopy of live cells. *Opt Express* 2009; **17**: 15585.

273 Doblas A, Roche E, Ampudia-Blasco FJ, Martínez-Corral M, Saavedra G, Garcia-Sucerquia J. Diabetes screening by telecentric digital holographic microscopy: DIABETES SCREENING BY TELECENTRIC DHM. *Journal of Microscopy* 2016; **261**: 285–290.

274 Javidi B, Markman A, Rawat S, O'Connor T, Anand A, Andemariam B. Sickle cell disease diagnosis based on spatio-temporal cell dynamics analysis using 3D printed shearing digital holographic microscopy. *Opt Express* 2018; **26**: 13614.

275 Moon I, Javidi B. Three-dimensional identification of stem cells by computational holographic imaging. *J R Soc Interface* 2007; **4**: 305–313.

276 Park Y, Diez-Silva M, Popescu G, Lykotrafitis G, Choi W, Feld MS *et al.* Refractive index maps and membrane dynamics of human red blood cells parasitized by Plasmodium falciparum. *Proceedings of the National Academy of Sciences* 2008; **105**: 13730–13735.

277 Jo Y, Park S, Jung J, Yoon J, Joo H, Kim M *et al.* Holographic deep learning for rapid optical screening of anthrax spores. *SCIENCE ADVANCES* 2017; : 10.

278 Dunn GA, Zicha D. Dynamics of fibroblast spreading. *Journal of Cell Science* 1995; **108**: 1239–1249.

279 Qiao H, Wu J. GPU-based deep convolutional neural network for tomographic phase microscopy with ℓ1 fitting and regularization. *J Biomed Opt* 2018; **23**: 1.

280 Barbastathis G, Ozcan A, Situ G. On the use of deep learning for computational imaging. *Optica* 2019; **6**: 921.

281 Shen D, Wu G, Suk H-I. Deep Learning in Medical Image Analysis. *Annu Rev Biomed Eng* 2017; **19**: 221–248.

282 Jiang S, Guo K, Liao J, Zheng G. Solving Fourier ptychographic imaging problems via neural network modeling and TensorFlow. *Biomed Opt Express* 2018; **9**: 3306.

283 Nguyen T, Xue Y, Li Y, Tian L, Nehmetallah G. Convolutional neural network for Fourier ptychography video reconstruction: learning temporal dynamics from spatial ensembles. ; : 24.

284 Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM *et al.* Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017; **542**: 115–118.

285 Rivenson Y, Zhang Y, Günaydın H, Teng D, Ozcan A. Phase recovery and holographic image reconstruction using deep learning in neural networks. *Light Sci Appl* 2018; **7**: 17141–17141.

286 Wang F, Bian Y, Wang H, Lyu M, Pedrini G, Osten W *et al.* Phase imaging with an untrained neural network. *Light Sci Appl* 2020; **9**: 77.

287 Liu T, de Haan K, Rivenson Y, Wei Z, Zeng X, Zhang Y *et al.* Deep learning-based super-resolution in coherent imaging systems. *Sci Rep* 2019; **9**: 3926.

288 Shi J, Chen Y, Zhang X. Broad-spectrum diffractive network via ensemble learning. *Opt Lett* 2022; **47**: 605.

289 Shi J, Wei D, Hu C, Chen M, Liu K, Luo J *et al.* Robust light beam diffractive shaping based on a kind of compact all-optical neural network. *Opt Express* 2021; **29**: 7084.

290 Chrabaszcz P, Loshchilov I, Hutter F. A Downsampled Variant of ImageNet as an Alternative to the CIFAR datasets. *arXiv:170708819 [cs]* 2017.http://arxiv.org/abs/1707.08819 (accessed 4 Jan2022).

291 Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing* 2004; **13**: 600–612.

292 Pawlowski E, Kuhlow B. Antireflection-coated diffractive optical elements fabricated by thin-film deposition. *Optical Engineering* 1994; **33**: 3537–3546.

293 Fluder G, Kowalik A, Rojek A, Sobczyk A, Choromański Z, Krężel J *et al.* Analysis of the influence of antireflective coatings on the diffraction efficiency of diffractive optical elements. *Opt Express* 2021; **29**: 13025.

294 Kulce O, Mengu D, Rivenson Y, Ozcan A. All-Optical Synthesis of an Arbitrary Linear Transformation Using Diffractive Surfaces. *arXiv:210809833 [physics]* 2021.http://arxiv.org/abs/2108.09833 (accessed 28 Aug2021).

295 Sakib Rahman MS, Ozcan A. Computer-Free, All-Optical Reconstruction of Holograms Using Diffractive Networks. *ACS Photonics* 2021; **8**: 3375–3384.

296 Huang Z, He Y, Wang P, Xiong W, Wu H, Liu J *et al.* Orbital angular momentum deep multiplexing holography via an optical diffractive neural network. *Opt Express* 2022; **30**: 5569.

297 Yan T, Wu J, Zhou T, Xie H, Xu F, Fan J *et al.* Fourier-space Diffractive Deep Neural Network. *Phys Rev Lett* 2019; **123**: 023901.

298 Luo X, Hu Y, Li X, Ou X, Lai J, Liu N. Metasurface-Enabled On-Chip Multiplexed Diffractive Neural Networks in the Visible. ; : 21.

299 Chen H, Feng J, Jiang M, Wang Y, Lin J, Tan J *et al.* Diffractive Deep Neural Networks at Visible Wavelengths. *Engineering* 2021; **7**: 1483–1491.

300 Seldowitz MA, Allebach JP, Sweeney DW. Synthesis of digital holograms by direct binary search. *Applied Optics* 1987; **26**: 2788.

301 Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. *arXiv:14126980 [cs]* 2017.http://arxiv.org/abs/1412.6980 (accessed 20 Jul2020).

302 Lin X, Rivenson Y, Yardimci NT, Veli M, Luo Y, Jarrahi M *et al.* All-optical machine learning using diffractive deep neural networks. *Science* 2018; **361**: 1004–1008.

303 Mengu D, Rivenson Y, Ozcan A. Scale-, Shift-, and Rotation-Invariant Diffractive Optical Networks. *ACS Photonics* 2020. doi:10.1021/acsphotonics.0c01583.

304 Qian C, Lin X, Lin X, Xu J, Sun Y, Li E *et al.* Performing optical logic operations by a diffractive neural network. *Light Sci Appl* 2020; **9**: 59.

305 Zhou T, Lin X, Wu J, Chen Y, Xie H, Li Y *et al.* Large-scale neuromorphic optoelectronic computing with a reconfigurable diffractive processing unit. *Nat Photonics* 2021; **15**: 367–373.

306 Rahman SS, Ozcan A. Computer-free, all-optical reconstruction of holograms using diffractive networks. ; : 19.

307 Luo Y, Zhao Y, Li J, Çetintaş E, Rivenson Y, Ozcan A. Computational Imaging Without a Computer: Seeing Through Random Diffusers at the Speed of Light. ; : 35.

308 Jiao S, Feng J, Gao Y, Lei T, Xie Z, Yuan X. Optical machine learning with incoherent light and a single-pixel detector. *Opt Lett* 2019; **44**: 5186.

309 Huang Z, Wang P, Liu J, Xiong W, He Y, Xiao J *et al.* All-Optical Signal Processing of Vortex Beams with Diffractive Deep Neural Networks. *Phys Rev Applied* 2021; **15**: 014037.

310 Shi J, Zhou L, Liu T, Hu C, Liu K, Luo J *et al.* Multiple-view D ^{2} NNs array: realizing robust 3D object recognition. *Opt Lett* 2021; **46**: 3388.

311 Ong JR, Ooi CC, Ang TYL, Lim ST, Png CE. Photonic Convolutional Neural Networks Using Integrated Diffractive Optics. *IEEE Journal of Selected Topics in Quantum Electronics* 2020; **26**: 1–8.

312 Shi J, Chen M, Wei D, Hu C, Luo J, Wang H *et al.* Anti-noise diffractive neural network for constructing an intelligent imaging detector array. *Opt Express* 2020; **28**: 37686.

313 Li Y, Chen R, Sensale-Rodriguez B, Gao W, Yu C. Real-time multi-task diffractive deep neural networks via hardware-software co-design. *Sci Rep* 2021; **11**: 11013.

314 Yan T, Wu J, Zhou T, Xie H, Xu F, Fan J *et al.* Fourier-space Diffractive Deep Neural Network. *Phys Rev Lett* 2019; **123**: 023901.

315 Colburn S, Chu Y, Shilzerman E, Majumdar A. Optical frontend for a convolutional neural network. *Appl Opt* 2019; **58**: 3179.

316 Elfadel IM, Wyatt JL Jr. The 'Softmax' Nonlinearity: Derivation Using Statistical Mechanics and Useful Properties as a Multiterminal Analog Circuit Element. In: Cowan J, Tesauro G, Alspector J (eds). *Advances in Neural Information Processing Systems*. Morgan-Kaufmann, 1994https://proceedings.neurips.cc/paper/1993/file/352407221afb776e3143e8a1a0577885-Paper.pdf.