Title
Machine Learning-Enabled Optical Sensors and Devices

Author
Veli, Muhammed

Publication Date
2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Machine Learning-Enabled Optical Sensors and Devices

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Electrical and Computer Engineering

by

Muhammed Veli

2020

ABSTRACT OF THE DISSERTATION

Machine Learning-Enabled Optical Sensors and Devices

by

Muhammed Veli

Doctor of Philosophy in Electrical and Computer Engineering

University of California, Los Angeles, 2020

Professor Aydogan Ozcan, Chair

Machine learning has been transforming many fields including optics by creating a new avenue for designing optical sensors and devices. This new paradigm takes a data driven approach, without focusing on underlying physics of the design. This new alternative and yet powerful method brings new advancements to traditional design tools and opens up numerous opportunities.

This dissertation introduces machine learning-enabled optical sensors and devices in which computational imaging and deep learning based design of devices tackle various challenges. First, a cost-effective and portable platform is presented to non-invasively detect and monitor a bacteria that resides in human ocular microbiome, *Staphylococcus aureus*. Contact lenses are designed to capture *S. aureus* using surface chemistry protocol, and sandwich immunoassay with polystyrene microbeads is performed to tag captured bacteria. Lens-free on-chip microscope is

used to obtain a single hologram of the contact lens surface and 3D surface of it is computationally reconstructed. Support vector machine based machine learning algorithm is employed to detect and count the amount of bacteria on contact lens surface. This platform, which only weighs 77 g, is controlled by laptop and provides ~16 bacteria/μL detection limit. This wearable sensor platform can be used to analyze and monitor other viruses and bacteria in tear with the appropriate modification to its surface chemistry protocol. Second, a novel physical mechanism is introduced, diffractive optical networks, to perform all-optical machine learning using passive diffractive layers that work together to implement various functions. This framework merges wave-optics with deep learning to all optically perform different tasks. A classification of handwritten digits and fashion products were demonstrated with 3D-printed diffractive optical networks. Moreover, a diffractive optical network is designed to function as an imaging lens in terahertz spectrum. This scalable platform can execute various functions at the speed of light with low power and help us to design exotic optical components. Third, terahertz pulse shaping architecture using diffractive optical surfaces is introduced. This platform engineers arbitrary broadband input pulse into desired waveform. Synthesis of various pulses has been demonstrated by designing and fabricating diffractive layers. This works constitutes the first demonstration of direct pulse shaping in terahertz spectrum with precise control of amplitude and phase of input broadband light over a wide frequency range. This approach can also find applications in other fields like optical communications, spectroscopy and ultra-fast imaging.

The dissertation of Muhammed Veli is approved.

Achuta Kadambi

Dino Di Carlo

Mona Jarrahi

Aydogan Ozcan, Committee Chair

University of California, Los Angeles

2020

To my mom, *Sevil*, and dad, *Elşen*

Table of Contents

List of Figures

List of Tables

Acknowledgements

First, I would like express to my sincere gratitude to Prof. Aydogan Ozcan. I think I am very lucky to work with him over the course of my PhD. His enthusiasm, mentorship and support were invaluable. It was very inspiring for my research to acquire his innovation, unique vision, creativity and unparalleled communication skills. I would like to also thank to Prof. Mona Jarrahi for giving opportunity to collaborate across disciplines and being in my committee. I also would like to thank to my doctoral committee members, Prof. Dino Di Carlo and Prof. Achuta Kadambi for their kind support during my PhD.

I would like thank to my colleagues and friends helping me thorough this journey: Euan Mcleod, Hatice Ceylan Koydemir, Deniz Mengu, Yichen Wu, Ani Ray, Hyou-arm Joung, Zoltan Gorocs, Zach Ballard, Calvin Brown, Mustafa Ugur Daloglu, Yi Luo, Nezih Tolga Yardimci, Alborz Feizi, Derek Tseng, Yibo Zhang, Hongda Wang, Artem Goncharov, Tianmin Shu, Wei Luo, Gokhan Senturk, and Robin Tuscher. I had an immense pleasure of working in such a collaborative and rich community of engineers and scientists.

I would like the thank American Chemical Society and American Association for the Advancement of Science, for allowing the permission to include a previous publication into a chapter.

I would like to acknowledge TUBITAK (Scientific and Technological Research Council of Turkey) for funding and their support during my PhD.

Last but not least, I would like to thank to my family for their great and unconditional support in my journey.

Vita

Muhammed Veli joined the Bio- and Nano-photonics Lab led by Prof. Aydogan Ozcan at UCLA in Fall 2014. He received Bachelor of Science in Electrical and Electronics Engineering from Bilkent University, Turkey in 2014 and Master of Science in Electrical Engineering from UCLA in 2016. He has co-authored 9 journal articles, and various conference proceedings on computational sensing and all-optical machine learning. He has been a Keynote Speaker in Biosensors 2018 and he is a winner of the Emil Wolf Outstanding Student Paper Competition in Frontiers in Optics, UCLA Graduate Division Fellowship, and Jean Bennett Memorial Student Travel Grant.

Selected Publications

1. **M. Veli**, D. Mengu, N. T Yardimci, Y. Luo, J. Li, Y. Rivenson, M. Jarrahi and A. Ozcan, Terahertz Pulse Shaping Using Diffractive Surfaces, Nature Communications, DOI: 10.1038/s41467-020-20268-z

2. J. Li, D. Mengu, N. T. Yardimci, Y. Luo, X. Li, **M. Veli**, Y. Rivenson, M. Jarrahi and A. Ozcan, Machine Vision using Diffractive Spectral Encoding, arXiv preprint arXiv:2005.11387

3. Y. Luo, D. Mengu, N.T. Yardimci, Y. Rivenson, **M. Veli**, M. Jarrahi and A. Ozcan, Design of task-specific optical systems using broadband diffractive neural networks, Light: Science & Applications (Nature Publishing Group) DOI: 10.1038/s41377-019-0223-1 (2019)

4. **M. Veli**, and A. Ozcan, Computational Sensing of Staphylococcus aureus on Contact Lenses using 3D Imaging of Curved Surfaces and Machine Learning, ACS Nano DOI: 10.1021/acsnano.7b08375 (2018)

5. X. Lin, Y. Rivenson, N.T. Yardimci, **M. Veli**, Y. Luo, M. Jarrahi and A. Ozcan, All-Optical Machine Learning Using Diffractive Deep Neural Networks, Science DOI: 10.1126/science.aat8084 (2018)

6. V. Mller, J.M. Sousa, H.C. Koydemir, **M. Veli**, D. Tseng, L. Cerqueira, A. Ozcan, N.F. Azevedo, F. Westerlund, Identification of Pathogenic Bacteria in Complex Samples Using a Smartphone Based Fluorescence Microscope, RSC Advances DOI: 10.1039/c8ra06473c (2018)

7. M.U. Daloglu, F. Lin, B. Chong, D. Chien, **M. Veli**, W. Luo, and A. Ozcan, 3D imaging of sex-sorted bovine spermatozoon locomotion, head spin and flagellum beating, Scientific Reports DOI:10.1038/s41598-018-34040-3 (2018)

8. E. McLeod, T.U. Dincer, **M. Veli**, Y.N. Ertas, C. Nguyen, W. Luo, A. Greenbaum, A. Feizi, and A. Ozcan, High-Throughput and Label-Free Single Nanoparticle Sizing Based on Time- Resolved On-Chip Microscopy, ACS Nano DOI: 10.1021/acsnano.5b00388 (2015)

9. E. McLeod, C. Nguyen, P. Huang, W. Luo, **M. Veli**, and A. Ozcan, Tunable vapor-condensed nanolenses, ACS Nano (2014) DOI:10.1021/nn502453h

Chapter 1 Machine Learning-Enabled Computational Sensing

Parts of this chapter have previously been published in M. Veli et al. “Computational Sensing of Staphylococcus aureus on Contact Lenses Using 3D Imaging of Curved Surfaces and Machine Learning”. ACS Nano, 2018, 12(3), 2554–2559.

In this chapter, I present a cost-effective and portable platform based on contact lenses for non-invasively detecting *Staphylococcus aureus*, which is part of the human ocular microbiome and resides on the cornea and conjunctiva. Using *Staphylococcus aureus*-specific antibodies and a surface chemistry protocol that is compatible with human tear, contact lenses are designed to specifically capture *Staphylococcus aureus.* After the bacteria capture on the lens, and right before its imaging, the captured bacteria are tagged with surface-functionalized polystyrene microparticles. These microbeads provide sufficient signal-to-noise ratio for the quantification of the captured bacteria on the contact lens, without any fluorescent labels, by 3D imaging of the curved surface of each lens using only one hologram taken with a lensfree on-chip microscope. After the 3D surface of the contact lens is computationally reconstructed using rotational field transformations and holographic digital focusing, a machine learning algorithm is employed to automatically count the number of beads on the lens surface, revealing the count of the captured bacteria. To demonstrate its proof-of-concept, we created a field-portable and cost-effective holographic microscope, which weighs 77 g, controlled by a laptop. Using daily contact lenses that are spiked with bacteria, we demonstrated that this computational sensing platform provides a detection limit of ~16 bacteria/μL. This contact lens–based wearable sensor can be broadly

applicable to detect various bacteria, viruses and analytes in tear using a cost-effective and portable computational imager that might be potentially used even at home by consumers.

1.1 Introduction

The human body is the host to several microorganisms, forming the human microbiota[1]. These microorganisms live symbiotically in various parts of the human body, including the conjunctiva[2], lungs[3], skin, saliva, gut, and vagina[4]. The relationship between the human microbiota and various diseases such as obesity, rheumatoid arthritis, and diabetes has drawn strong attention to the analysis of the human microbiome[5], also motivated by personalized treatments and medicine. Several microorganisms are also found in the ocular microbiota, one of which is *Staphylococcus aureus*. It is a gram-positive bacterium and is present in different parts of the body, including the conjunctiva, nose, and skin[6–8]. It is colonized in approximately 30% of humans[9]. It is also a human pathogen, and the carriers of *Staphylococcus aureus* have a higher risk of infection. It causes a variety of diseases like pneumonia, endocarditis, arthritis, and bacteremia[6].

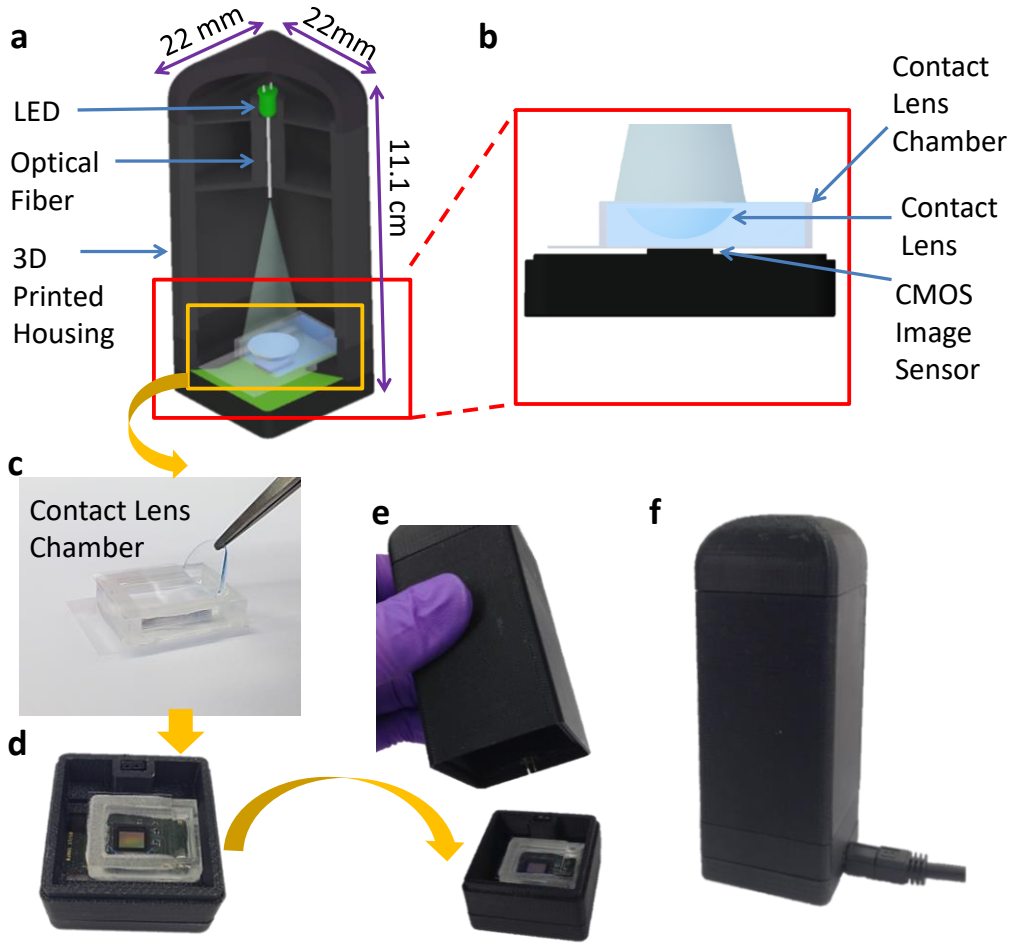

Fig. 1.1 Schematic illustration of device and workflow (a) 3D schematic illustration and (b) a cross section of the portable lensfree microscope including a contact lens chamber. (c) Contact lens insertion into a PBS filled contact lens chamber. (d) Placement of the chamber on a CMOS image sensor. (e, f) Photographs of the lensfree microscope.

Monitoring of the human microbiome with the current methods is relatively costly and time-consuming. 16S rRNA and whole-genome shotgun metagenomics gene-based sequencing are among the most commonly used techniques for the analysis of microbiome[10,11]. 16S rRNA sequencing starts with the extraction of the DNA from the isolated sample, and it is followed by

polymerase chain reaction (PCR) amplification, cycle sequencing, and database comparison[12]. As an additional challenge for monitoring of ocular microbiota, collecting samples from the ocular surface is quite tedious. It involves conjunctival swabs or tear fluid collection[8,13] which require laborious sample collection steps as well as relatively expensive and bulky laboratory equipment for the sample analysis.

Here, we present a cost-effective and portable platform that is based on contact lenses for detecting and monitoring *Staphylococcus aureus*, which is found in the human ocular microbiome. In this study, the detection of *Staphylococcus aureus* is enabled by a surface functionalized contact lens[14], a lens-free computational imaging setup[15–18], and a machine learning-based algorithm to quantify the amount of bacteria captured on the contact lens. The surface of each contact lens is functionalized by a layer-by-layer (LBL) coating technique[19] which is shown to be compatible with the human tear. The LBL coating technique creates biofunctional films on different surface morphologies, and with the appropriate coating material, it enables contact lenses to be worn without any damage to the cornea[14]. Seven layers of poly(styrene sulfonate) (PSS) and poly(allylamine hydrochloride) (PAH) coating on a contact lens create a stable structure for binding of the anti-*Staphylococcus aureus* antibody to the surface of the contact lens. We used antibody attached daily contact lenses for specifically capturing *Staphylococcus aureus* particles that normally reside on the human cornea and conjunctiva. These surface functionalized contact lenses are expected to be worn for e.g., 12–16 h for continuously capturing *Staphylococcus aureus* in tear. In this work, however, we spiked daily contact lenses with bacteria and after the capture process, and right before the 3D imaging of the lens surface, we attached 5-μm polystyrene beads to specifically bind to *Staphylococcus aureus* particles already captured on the lens. *Staphylococcus aureus* has a diameter of 0.75–1

μm[20], and therefore 5-μm beads make their detection and counting easier using lensfree imaging of the contact lens 3D surface with a ***single*** hologram.

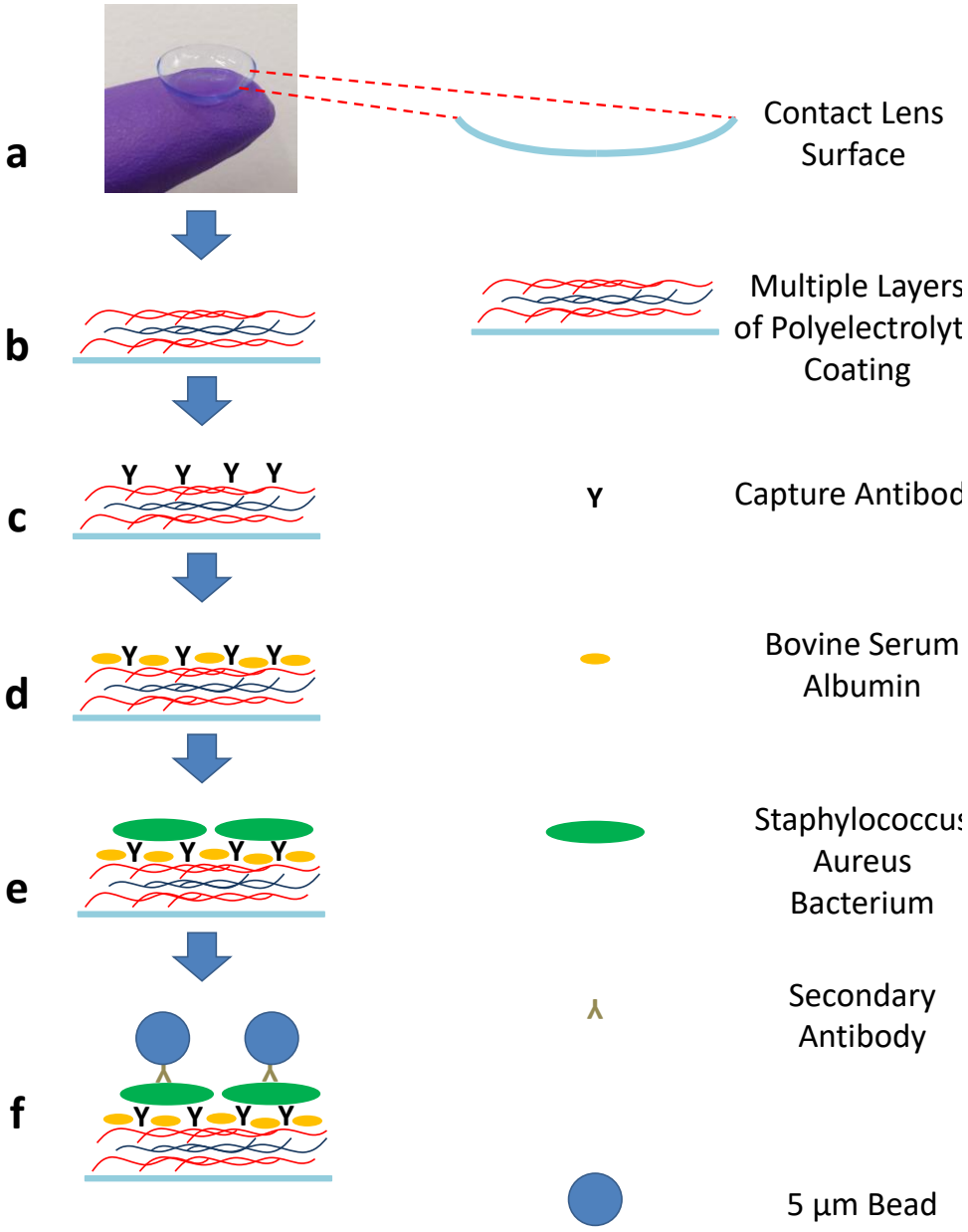

Fig. 1.2 Schematic illustration of the contact lens surface functionalization and bead-based immunoassay steps

In fact, microscopic imaging of a contact lens is very challenging using traditional microscopy tools because it is large and has a curved 3D structure. Conventional microscopes have a small field of view (FOV) and a limited depth of field, making it extremely difficult and time-consuming to image the 3D surface of a contact lens using traditional imaging approaches. Our computational imaging platform is not only cost-effective and field-portable, but also offers a very large FOV (~30 mm^2) and depth of field (> 1cm), which make it ideal for label-free imaging of the 3D surface of a contact lens using a single hologram, without the need for any mechanical scanning or fine alignments (Figure 1.1). Our lens-free on-chip imaging setup utilizes a multimode-fiber-coupled light emitting diode (LED), which emits partially coherent light at 527 nm. In order to image the contact lenses with their natural curvy shape, they are placed in a custom-designed chamber filled with a phosphate-buffered saline (PBS) solution. An inexpensive complementary metal oxide semiconductor (CMOS) image sensor captures a single hologram of each contact lens, which is then reconstructed to automatically identify the captured micro-beads on the lens surface. This analysis includes a tilt correction algorithm that handles the image distortions caused by the curvature of the contact lenses, a holographic 3D image reconstruction algorithm, as well as a support vector machine (SVM) based learning model that can specifically detect the captured 5-μm beads and digitally separate them from other unwanted particles deposited on the lens surface.

Using daily contact lenses that are surface functionalized and spiked with bacteria, we demonstrated a detection limit of ~16 bacteria/μL with this computational 3D imaging and machine learning approach. The presented wearable sensing framework using surface functionalized contact lenses and computational imaging can be broadly applicable to sense

numerous other target bacteria, viruses and analytes in tear samples and might even be suitable for consumer use at home.

1.2 Results

Acuvue Moist 1-day contact lenses have an inherently ionic surface structure. Etafilcon A of these contact lenses is negatively charged[21] at the physiological pH so that the innermost layer of the positively charged polyelectrolyte, PAH, can bind to the contact lens surface electrostatically. In a similar fashion, the chemical functionalization of the contact lens surface was performed with an alternative LBL deposition of the positively charged polyelectrolyte, PAH, and the negatively charged polyelectrolyte, PSS, by using the electrostatic interaction between them. The outermost layer was coated with the positively charged PAH to enable the electrostatic interaction between the functionalized contact lens surface and the anti-*Staphylococcus aureus* antibody, which is negatively charged at the physiological pH. These seven alternating layers of coating on the contact lens surface created a stable interface for the successive antibody binding and bead-based immunoassay formation (Figure 1.2). Importantly, previous work showed that this polyelectrolyte coating did not exhibit cell toxicity[14].

As detailed in the Methods section, the digital 3D surface mapping of each contact lens under test provides the rotation angle for each sub-region on the lens surface that we need for field rotations. Figure 1.3-d shows the digitally reconstructed 3D surface of a contact lens. The color map indicates the distance of each point on the contact lens surface from the CMOS image sensor plane. In our experiments, we observed a maximum tilt of approximately 20° at the corners of the contact lens with respect to the CMOS sensor. After the tilt correction of each sub-region of interest, we create a set of images, each of which is tangential to the contact lens

surface. By digitally stitching all these tilt-corrected small tiles, we obtained an in-focus image of the contact lens surface, covering approximately 31 mm^2 on the lens surface, which is larger than the active area of the CMOS imager, as expected. In Figure 1.3-e, the reconstructions of some of these sub-regions are shown without any tilt correction and after tilt correction. We can clearly observe that the 5-μm beads are in focus, as desired, over the entire FOV after the tilt correction, whereas in the regular reconstructions, they are out of focus at the edges of each region of interest. This phenomenon is more obvious if one looks at the corners of the contact lens; for example, in regions 1 and 2, the impact of tilt corrections is more apparent compared to region 3 (Figure 1.3).

To explore the detection limit of our platform, five solutions of *Staphylococcus aureus* at different concentrations were applied after the surfaces of the contact lenses were functionalized as described in the Methods section. At each concentration of bacteria, we used three individual contact lenses to test the repeatability of our approach. After the incubation of the streptavidin-coated antibody conjugated 5-μm beads, the contact lenses were imaged using our lens-free on-chip microscope, and the 5-μm beads were automatically counted using an SVM-based algorithm described in the Methods section. Figure 1.4 shows the number of the detected beads on contact lens surfaces for different concentrations of *Staphylococcus aureus*. A detection limit of 16.3 bacteria/μL was achieved using our platform, based on $\mu+3\sigma$ of our control samples, where μ refers to the mean and σ is the standard deviation. Previous results reported that the human eye can contain increased concentrations of bacteria (e.g., >10 cfu/μL) even for asymptomatic individuals.[8] Considering the fact that under a bacterial infection the concentration of bacteria in tear is expected to be significantly higher than this baseline, we believe that our sensing limit is relevant for the detection of such infections. A further improvement in our

sensing performance can potentially be achieved by increasing our training data size and the number of the spatial features used in the SVM-based particle detection algorithm, which can help us reduce the rate of false positives. Convolutional neural network based deep learning approaches could also be utilized to further advance our results, which is left as future work.

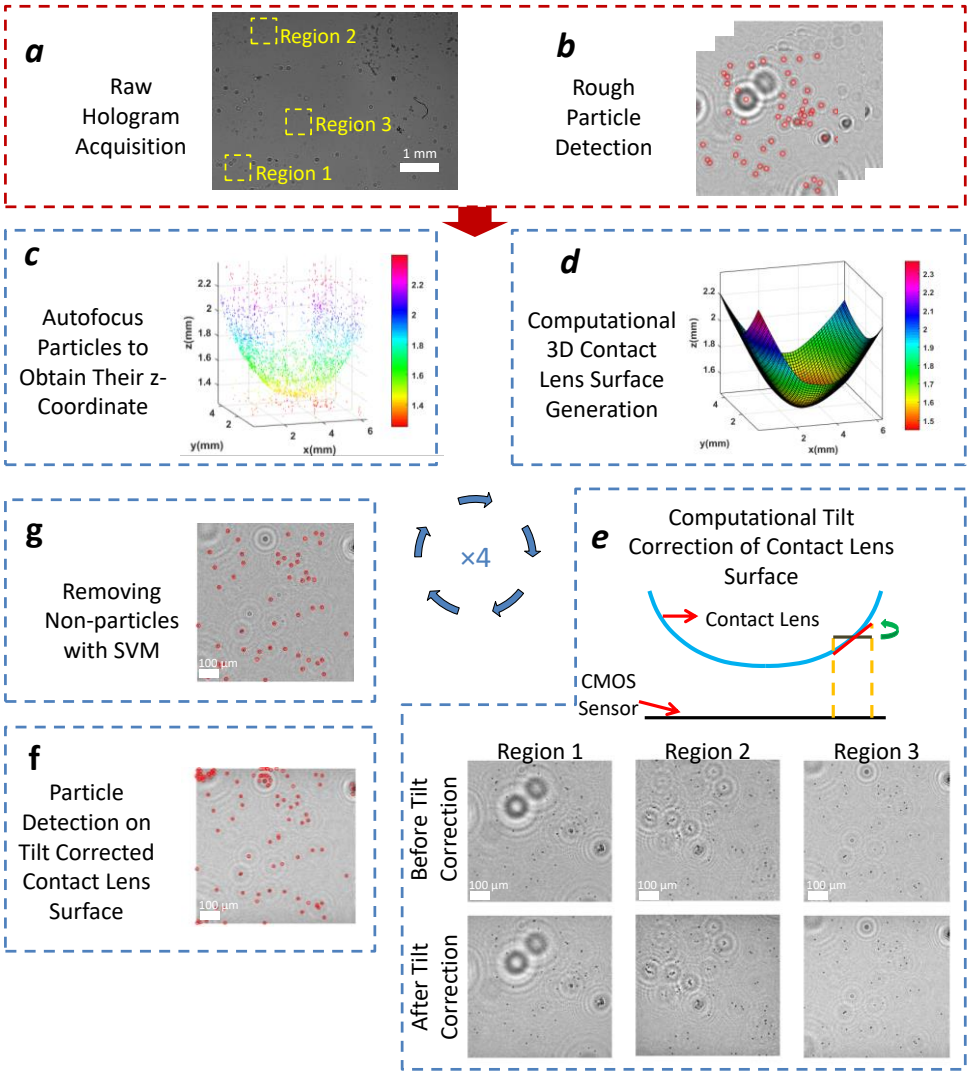

Fig. 1.3 Schematic diagram of the image reconstruction and processing steps

1.3 Discussion

The presented work makes use of contact lenses as the basis of a computational sensing platform. Although not demonstrated in this work since we did not experiment with human samples, we believe that this wearable method is more convenient for patients when compared to e.g., conjunctival swabs and tear collection. We also believe that this method should increase the efficiency of sample collection because the contact lens will be worn throughout the day for e.g., 12–16 h and it will be continuously in contact with the human tear.

We demonstrated a proof-of-concept platform that can potentially monitor and non-invasively analyze the human ocular microbiome in a cost-effective manner. This platform can automatically detect and enumerate *Staphylococcus aureus* particles that are captured on contact lenses. It includes a field-portable lens-free microscope, a custom-made contact lens chamber, surface functionalized contact lenses, and an automated holographic image reconstruction and processing algorithm that also utilizes machine learning. We have achieved a detection limit of 16.3 CFU/μL. We believe that this platform might serve as a promising tool for the analysis and monitoring of the human ocular microbiome and can be broadly applicable to other target bacteria, viruses and analytes that can be sensed using wearable and flexible substrates, including but not limited to contact lenses.

1.4 Materials and Methods

The procedure for the automated detection and quantification of *Staphylococcus aureus* starts with the functionalization of the surface of a contact lens, and it is followed by a bead-based immunoassay on the lens surface to achieve specificity and sensitivity. After the capture of the target bacteria, each contact lens under test is placed in a custom-designed sample holder, and its

lens-free hologram is taken using our portable on-chip microscope. Each hologram is then rapidly reconstructed to reveal a microscopic image of the 3D surface of the contact lens, which is then analyzed using an automated machine-learning algorithm to estimate the count of *Staphylococcus aureus* captured on the lens surface.

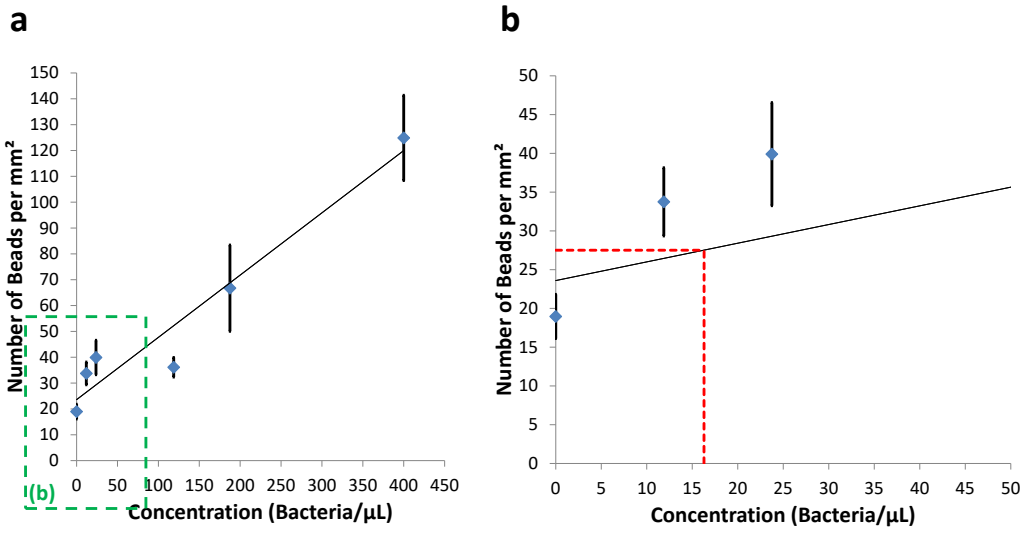

Fig. 1.4 Limit of detection of the platform (a) The number of the detected beads on the contact lens surface per mm² as a function of the bacteria concentration. (b) Zoomed in version of (a), red dashed line refers to µ+3σ of the control samples.

Materials

Poly(sodium 4-styrenesulfonate) (PSS; M_w ~200000 g mol^{-1}) (product no. 561967), bovine serum albumin (BSA) (product no. B4287), phosphate buffered saline (PBS pH 7.4) (product no. P3813), and Tween®20 (product no. P9416) were purchased from Sigma–Aldrich. Poly(allylamine hydrochloride) (PAH; M_w ~120000–200000 g mol^{-1}) (product no. 43092) was purchased from Alfa Aesar. Anti-*Staphylococcus aureus* antibody (product no. ab73962) and

Anti-*Staphylococcus aureus* antibody biotin (product no. ab35192) were purchased from Abcam. Contact lenses (Acuvue Moist 1-day) were purchased from Contact Lenses Canada. *Staphylococcus aureus* bacteria (product no. 27660) were purchased from American Type Culture Collection (ATCC). Sodium chloride (NaCl) (product no. SX0420) was purchased from Millipore Sigma. Streptavidin coated 5-µm polystyrene beads (product no. SVP-50-5) were purchased from Spherotech Inc. Acrylic glass (PMMA) (product no. 8560K354) was purchased from McMaster-Carr. The LED (product no. C503B-GAN-CB0F0791-ND) was purchased from Digikey and the multimode fiber (product no. FG105LCA) was purchased from Thorlabs. The coverslips were purchased from Fisher Scientific. Reagent grade water was used throughout the experiments.

Design of the portable lens-free on-chip microscope

Our lens-free microscope consists of an LED that emits green light with a peak wavelength of 527 nm, a multimode fiber (105 µm core diameter), a CMOS image sensor with a pixel size of 1.67 µm (MT9J003STM/STC, ON Semiconductor), a custom-made contact lens holder, and a 3D printed housing that holds all the components (Figure 1.1). The design of the housing was done using Autodesk Inventor Professional, and it was printed using a 3D printer and acrylonitrile butadiene styrene (ABS) material (Dimensions Elite, Stratasys).

The top part of the contact lens chamber is a 22x22 mm No.1 coverslip (thickness: 120 µm), and the bottom part is formed by a 24x35 mm No.0 coverslip (thickness of 70 µm). The rectangular sidewalls of the chamber were prepared using a laser cutting device with PMMA (thickness of 6 mm). The coverslips and PMMA were glued together using an epoxy (Figure 1.1-c). After the bead-based immunoassay, each contact lens was placed inside the PBS-filled

contact lens chamber for holographic imaging. This sample holder provides an environment for the contact lenses to retain their natural 3D structure and curvature while enabling us to image them by preventing surface fluctuations that may occur at the liquid–air interface. The holder was then placed directly on the top of the CMOS image sensor. The fiber-coupled LED provided sufficient spatial coherence to capture a single hologram that contains all the microparticles captured over the 3D surface of the contact lens. The exposure time of each hologram was ~50 ms, which is short enough to avoid any problems related to a potential shift in the contact lens position within the sample holder. For illumination, a hole was drilled on the LED package, and a multi-mode fiber was inserted into that hole and fixed with an optical glue[22]. This fiber-coupled LED was then placed approximately 5 cm away from the CMOS sensor plane. The LED was powered by the CMOS sensor board.

Contact lens surface functionalization

After unpacking the daily contact lenses, we washed them by dipping each one of the lenses into a PBS solution (10 mM, pH 7.4) to equilibrate the polymer surface of the contact lenses. The surface was then functionalized using the LBL deposition of polyelectrolytes, PAH, and PSS. First, the contact lenses were dipped into cationic polyelectrolyte, PAH (5 mg mL^{-1} in 0.5 M NaCl) for 15 min. Then, they were washed with 0.5 M NaCl solution three times. Next, the contact lenses were dipped into anionic polyelectrolyte, PSS (5 mg mL^{-1}, in 0.5 M NaCl), for 15 min. Then, they were washed again with 0.5 M NaCl solution three times. The same PAH and PSS deposition steps were repeated until seven layers of polyelectrolyte coating (i.e., PAH-PSS-PAH-PSS-PAH-PSS-PAH) were created on the surface of each contact lens (Figure 1.2-b).

Bead-based immunoassay and related experimental procedures

200 μL of 10 μg/mL captured antibody anti-*Staphylococcus aureus* solution in 10 mM PBS (pH 7.4) was incubated on the contact lenses, which were previously coated with polyelectrolytes as detailed earlier (Figure 1.2-c). After this incubation for 16 h, they were washed with 10 mM PBS at pH 7.4 to remove the excess antibody from the contact lenses. Then, the lens surface was blocked by a buffer (1% BSA in 10 mM PBS), incubated for 2 h to reduce the non-specific binding of the antibody attached polystyrene beads (Figure 1.2-d). The contact lenses were then washed once with 10 mM PBS at pH 7.4.

To mimic the daily use of the contact lenses, 200 μL solution of 4% (v/v) formaldehyde fixed *Staphylococcus aureus* (at different concentrations) in 10 mM PBS (pH 7.4) was incubated on the contact lenses for 16 h (Figure 1.2-e). During this incubation period, *Staphylococcus aureus* would adhere to the antibody coated contact lens surface. After this incubation, the contact lenses were washed once with 10 mM PBS (pH 7.4). Then, at the testing phase (i.e., after each contact lens captured the *Staphylococcus aureus* particles on its surface) the 200-μL bead and antibody mixture was incubated on the contact lenses for 2 h (Figure 1.2-f). To prepare these streptavidin-coated 5-μm beads, they were mixed with 27 μg/mL biotin-anti-*Staphylococcus aureus* in a separate tube in 10 mM PBS (pH 7.4) for 1 h. The strong interaction between streptavidin and biotin enables antibody-conjugated microbeads to specifically attach to *Staphylococcus aureus*. This approach provided specificity, helping us correlate the number of bacteria with the number of microbeads. Finally, the contact lenses were washed thrice with a washing buffer (10 mM PBS, 0.1% BSA, and 0.05% Tween ® 20 at pH 7.4) to remove the excess beads before their imaging using the holographic on-chip microscope.

Automated analysis of the holograms of contact lenses

We prepared an image processing algorithm (Figure 1.3) for the automated detection and counting of *Staphylococcus aureus* captured on contact lenses. This algorithm starts with the holographic reconstruction of the 3D contact lens surface and uses an SVM-based learning algorithm for label-free classification of the captured particles on the lens surface. First, a rough estimate of all the particles on a contact lens was made by reconstructing its lens-free hologram at all the possible object planes with a vertical spacing of 5 μm as shown in Figure 1.3-b. The possible 5-μm bead candidates were then detected from each reconstructed amplitude image by a simple threshold. In order to have a better accuracy in the axial position estimation of each particle, an autofocusing algorithm using the Tamura coefficient was employed on these potential particle candidates[23–27]. The resulting x-y-z positions of these beads were then selected as the sampled points on the contact lens 3D surface (Figure 1.3-c). A physical constraint based on this initially estimated 3D shape of the contact lens was also applied to remove possible detection artifacts and false positives due to unbound floating particles in the sample holder. A locally-weighted linear regression was then performed on these points to digitally generate a 3D map of each contact lens surface that is under test, as shown in Figure 1.3-d. This step is important to properly image and count the 5-μm beads captured on the surface of the contact lenses, and that is why we first reconstructed the 3D contact lens surface digitally, and then employed a tilt correction algorithm[8,29] to smaller regions of interest on the contact lens surface to digitally bring each region of interest in focus. There are different approaches to image tilted objects such as extended-focus imaging (EFI) [30,31] and rotational field transformations[28,29]. In this work, we used the latter approach, which is based on the angular spectrum method. This tilt correction algorithm involves two Fast Fourier Transform (FFT) operations and one interpolation operation. In essence, with the knowledge of the complex optical field at one of the planes, we

can obtain the complex field at another rotated plane by using this tilt correction algorithm. In order to perform this operation, each lens-free hologram of a contact lens was digitally divided into 192 equal tiles (i.e., smaller regions of interest) and each tile was processed separately. These tiles were reconstructed at their corresponding heights, where their centers intersected with the computationally generated 3D contact lens surface. The natural curvature of the contact lenses prevents each tile to be completely in-focus, making the edges of the tiles out of focus. To mitigate this, the reconstructed tiles were rotated using the tilt correction algorithm so that they became tangent to the computationally generated contact lens 3D surface (Figure 1.3-e). For each tile, we calculated the rotation angle based on the reconstructed 3D contact lens surface.

To further eliminate false positives, we employed an SVM based learning algorithm[26] to distinguish 5-μm beads from other non-specifically bound particles on the lens surface that appear in the reconstructed and tilt-corrected holographic images. Each object's size, intensity and ratio of the Tamura coefficient at the focus plane with respect to four other planes were fed into an SVM algorithm forming a set of *six* features. The SVM was then trained on approximately 3,000 particles, where 1,200 of them were 5-μm beads, all of which were manually labeled, forming our training data. Figure 1.3-g shows the blind detection of the target beads captured on the contact lens surface after the SVM was successfully trained. The steps shown in Figure 1.3-c to Figure 1.3-g were repeated three more times for refining the detection results.

Chapter 2 All-optical Machine Learning Using Diffractive Networks

Parts of this chapter have previously been published in X. Lin et al. “All-optical machine learning using diffractive deep neural networks” Science, 361, 1004-1008 (2018). In this chapter, all-optical machine learning platform is introduced.

Deep learning has been transforming our ability to execute advanced inference tasks using computers. Here we introduce a physical mechanism to perform machine learning by demonstrating an all-optical diffractive deep neural network (D^2NN) architecture that can implement various functions following the deep learning–based design of passive diffractive layers that work collectively. We created 3D-printed D^2NN s that implement classification of images of handwritten digits and fashion products, as well as the function of an imaging lens at a terahertz spectrum. Our all-optical deep learning framework can perform, at the speed of light, various complex functions that computer-based neural networks can execute; will find applications in all-optical image analysis, feature detection, and object classification; and will also enable new camera designs and optical components that perform distinctive tasks using D^2NN s.

2.1 Introduction

Deep learning is one of the fastest-growing machine learning methods[32]. This approach uses multilayered artificial neural networks implemented in a computer to digitally learn data representation and abstraction and to perform advanced tasks in a manner comparable or even superior to the performance of human experts. Recent examples in which deep learning has made major advances in machine learning include medical image analysis[33], speech recognition[34],

language translation[35], and image classification[36], among others[32,37]. Beyond some of these mainstream applications, deep learning methods are also being used to solve inverse imaging problems[38–44].

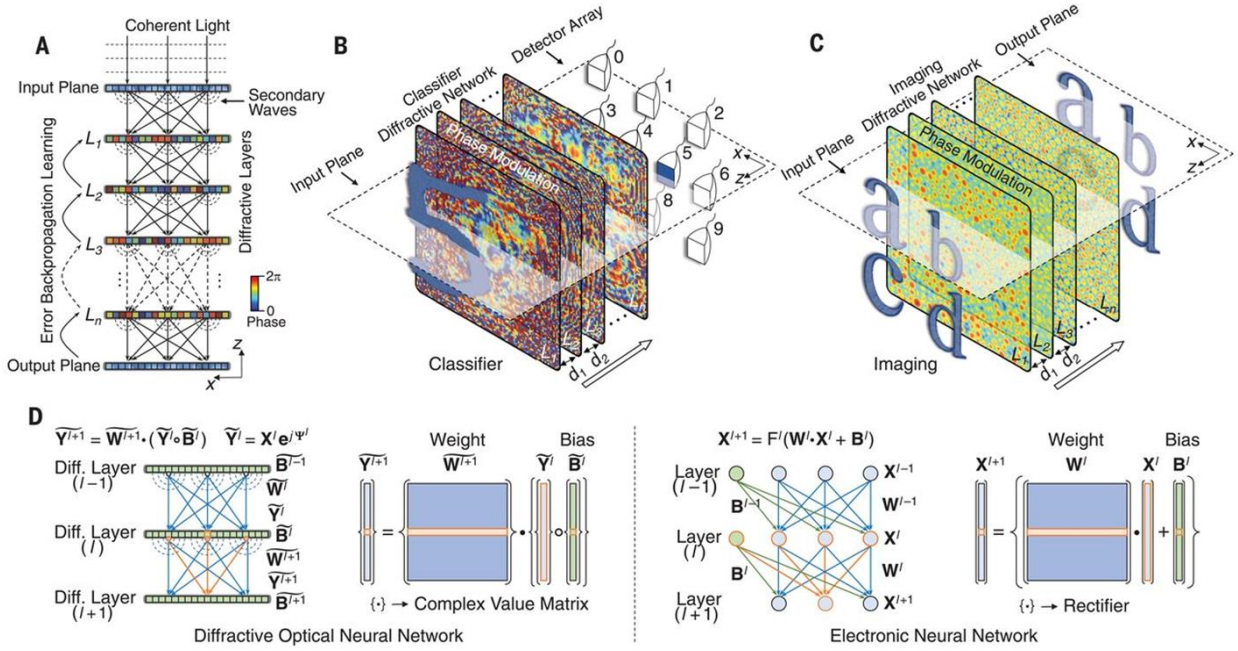

Fig. 2.1 Diffractive deep neural networks (D^2NNs). (**A**) A D^2NN comprises multiple transmissive (or reflective) layers, where each point on a given layer acts as a neuron, with a complex-valued transmission (or reflection) coefficient. The transmission or reflection coefficients of each layer can be trained by using deep learning to perform a function between the input and output planes of the network. After this learning phase, the D^2NN design is fixed; once fabricated or 3D-printed, it performs the learned function at the speed of light. L, layer. (**B** and **C**) We trained and experimentally implemented different types of D^2NNs: (B) classifier (for handwritten digits and fashion products) and (**C**) imager. d, distance. (**D**) Comparison between a D^2NN and a conventional neural network (14). Based on coherent waves, the D^2NN operates on complex-valued inputs, with multiplicative bias terms. Weights in a D^2NN are based on free-space diffraction and determine the interference of the secondary waves that are phase- and/or amplitude-modulated by the previous layers. "**o**" denotes a Hadamard product operation. "Electronic neural network" refers to the conventional neural network virtually implemented in a computer. **Y**, optical field at a given

layer; Ψ, phase of the optical field; **X**, amplitude of the optical field; F, nonlinear rectifier function [see 2.4 Materials and Methods for a discussion of optical nonlinearity in D^2NN].

Here we introduce an all-optical deep learning framework in which the neural network is physically formed by multiple layers of diffractive surfaces that work in collaboration to optically perform an arbitrary function that the network can statistically learn. Whereas the inference and prediction mechanism of the physical network is all optical, the learning part that leads to its design is done through a computer. We term this framework a diffractive deep neural network (D^2NN) and demonstrate its inference capabilities through both simulations and experiments. Our D^2NN can be physically created by using several transmissive and/or reflective layers, where each point on a given layer either transmits or reflects the incoming wave, representing an artificial neuron that is connected to other neurons of the following layers through optical diffraction (Fig. 2.1A). In accordance with the Huygens-Fresnel principle, our terminology is based on each point on a given layer acting as a secondary source of a wave, the amplitude and phase of which are determined by the product of the input wave and the complex-valued transmission or reflection coefficient at that point. Therefore, an artificial neuron in a D^2NN is connected to other neurons of the following layer through a secondary wave modulated in amplitude and phase by both the input interference pattern created by the earlier layers and the local transmission or reflection coefficient at that point. As an analogy to standard deep neural networks (Fig. 2.1D), one can consider the transmission or reflection coefficient of each point or neuron as a multiplicative "bias" term, which is a learnable network parameter that is iteratively adjusted during the training process of the diffractive network, using an error back-propagation method. After this numerical training phase, the D^2NN design is fixed and the transmission or

reflection coefficients of the neurons of all layers are determined. This D^2NN design, once

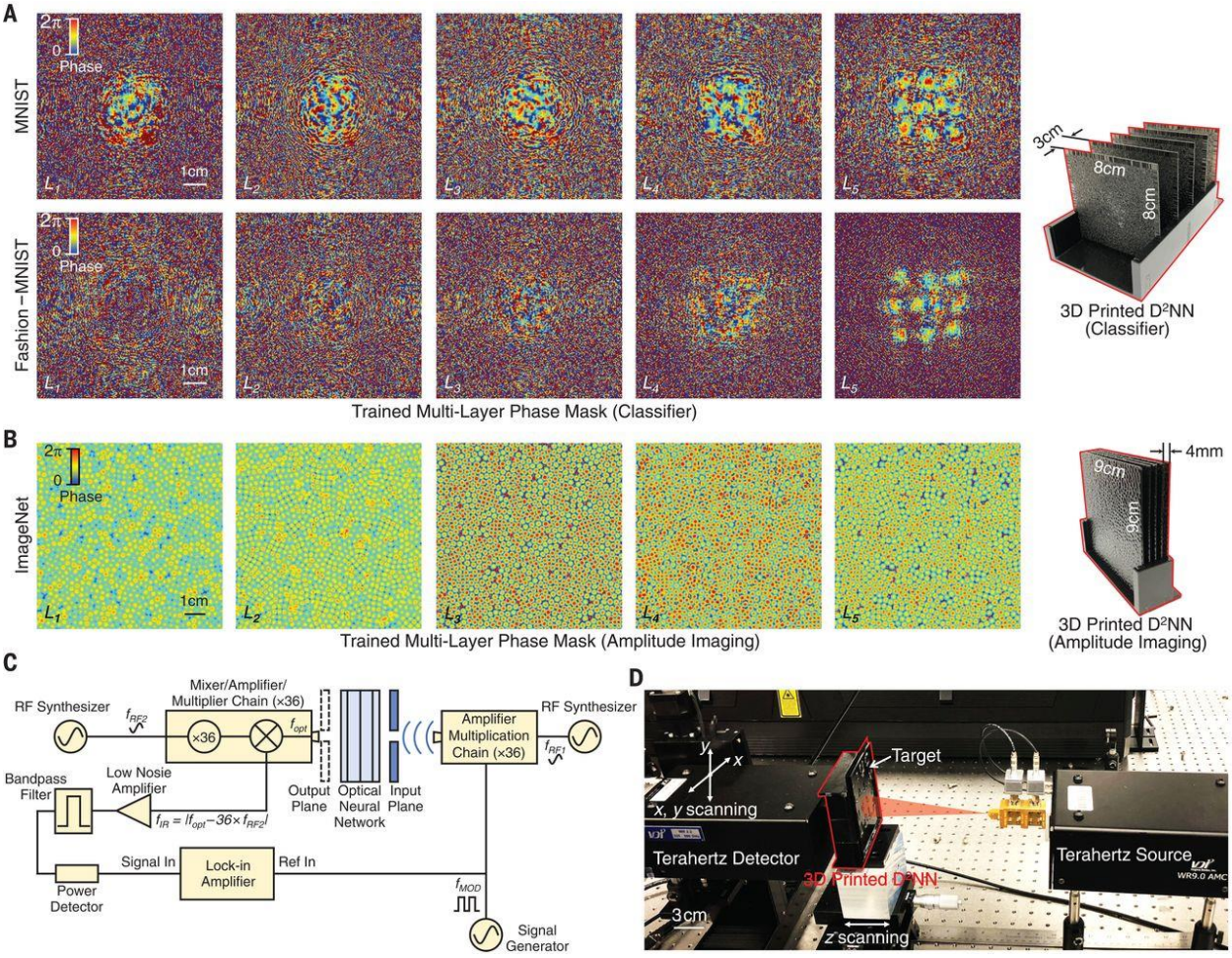

Fig. 2.2 Experimental testing of 3D-printed D^2NNs. (**A** and **B**) After the training phase, the final designs of five different layers (L1, L2, …, L5) of the handwritten digit classifier, fashion product classifier, and the imager D^2NNs are shown. To the right of the network layers, an illustration of the corresponding 3D-printed D2NN is shown. (**C** and **D**) Schematic (C) and photo (D) of the experimental terahertz setup. An amplifier-multiplier chain was used to generate continuous-wave radiation at 0.4 THz, and a mixer-amplifier-multiplier chain was used for the detection at the output plane of the network. RF, radio frequency; f, frequency.

physically fabricated using techniques such as 3D-printing or lithography, can then perform, at the speed of light, the specific task for which it is trained, using only optical diffraction and

passive optical components or layers that do not need power, thereby creating an efficient and fast way of implementing machine learning tasks.

2.2 Results

In general, the phase and amplitude of each neuron can be learnable parameters, providing a complex-valued modulation at each layer, which improves the inference performance of the diffractive network. For coherent transmissive networks with phase-only modulation, each layer can be approximated as a thin optical element (Fig. 2.1). Through deep learning, the phase values of the neurons of each layer of the diffractive network are iteratively adjusted (trained) to perform a specific function by feeding training data at the input layer and then computing the network's output through optical diffraction. On the basis of the calculated error with respect to the target output, determined by the desired function, the network structure and its neuron phase values are optimized via an error back-propagation algorithm, which is based on the stochastic gradient descent approach used in conventional deep learning.

To demonstrate the performance of the D^2NN framework, we first trained it as a digit classifier to perform automated classification of handwritten digits, from 0 to 9 (Figs. 2.1B and 2.2A). For this task, phase-only transmission masks were designed by training a five-layer D^2NN with 55,000 images (5000 validation images) from the MNIST (Modified National Institute of Standards and Technology) handwritten digit database[45]. Input digits were encoded into the amplitude of the input field to the D^2NN, and the diffractive network was trained to map input digits into 10 detector regions, one for each digit. The classification criterion was to find the detector with the maximum optical signal, and this was also used as a loss function during the network training.

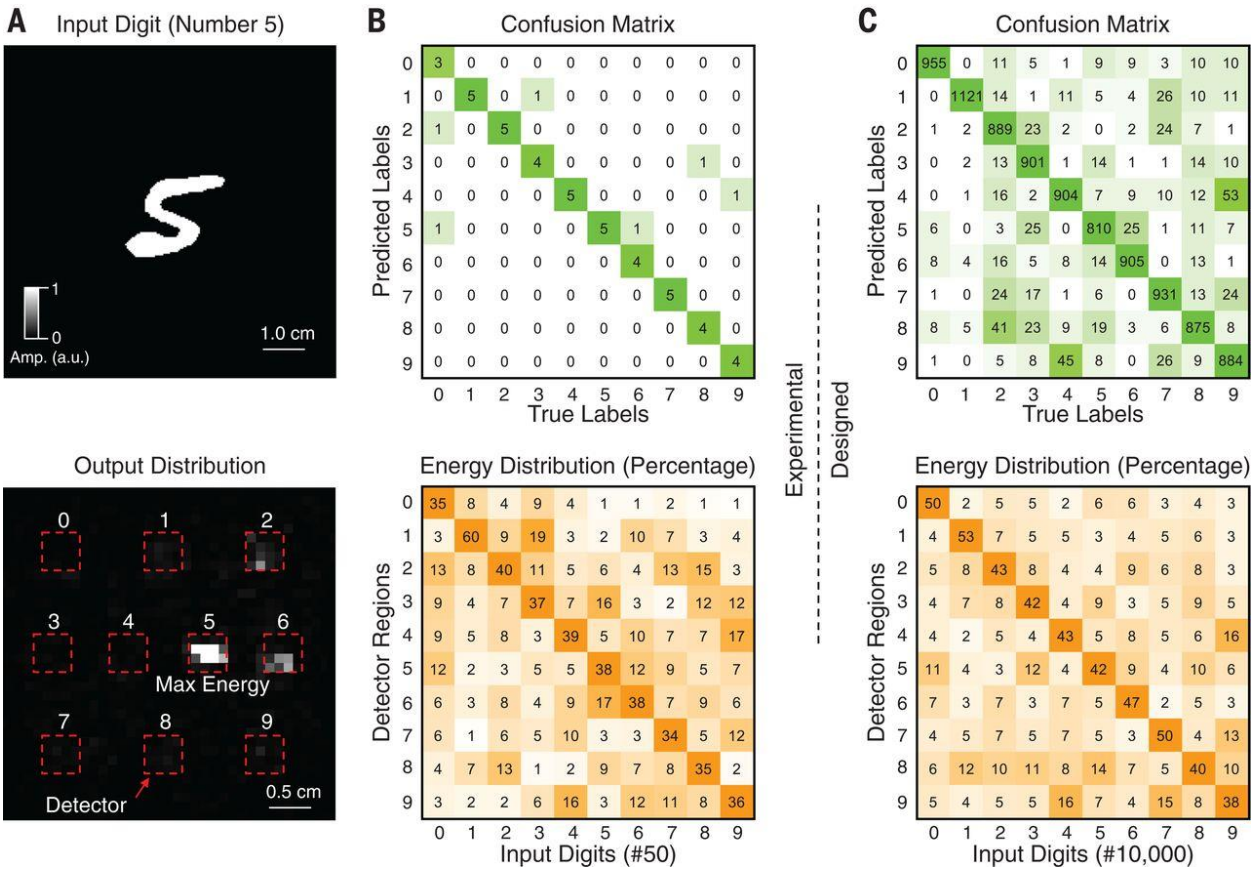

Fig. 2.3 Handwritten digit classifier D^{2}NN. (A) A 3D-printed D^{2}NN successfully classifies handwritten input digits (0, 1, …, 9) on the basis of 10 different detector regions at the output plane of the network, each corresponding to one digit. As an example, the output image of the 3D-printed D^{2}NN for a handwritten input of "5" is demonstrated, where the red dashed squares represent the trained detector regions for each digit. Other examples of our experimental results are shown in fig. 2.13. (B) Confusion matrix and energy distribution percentage for our experimental results, using 50 different handwritten digits (five for each digit) that were 3D-printed, selected among the images for which numerical testing was successful. (C) Same as (B), except summarizing our numerical testing results for 10,000 different handwritten digits (~1000 for each digit), achieving a classification accuracy of 91.75% using a five-layer design. Our classification accuracy increased to 93.39% by increasing the number of diffractive layers to seven, using a patch of two additional diffractive layers added to an existing and fixed D2NN (fig. 2.6).

After training, the design of the D^{2}NN digit classifier was numerically tested using 10,000 images from the MNIST test dataset (which were not used as part of the training or validation

image sets) and achieved a classification accuracy of 91.75% (Fig. 2.3C). In addition to the classification performance of the diffractive network, we also analyzed the energy distribution observed at the network output plane for the same 10,000 test digits (Fig. 2.3C), the results of which clearly demonstrate that the diffractive network learned to focus the input energy of each handwritten digit into the correct (i.e., the target) detector region, in accord with its training. With the use of complex-valued modulation and increasing numbers of layers, neurons, and connections in the diffractive network, our classification accuracy can be further improved (figs. 2.5 and 2.6). For example, fig. 2.6 demonstrates a Lego-like physical transfer learning behavior for D^2NN framework, where the inference performance of an already existing D^2NN can be further improved by adding new diffractive layers—or, in some cases, by peeling off (i.e., discarding) some of the existing layers—where the new layers to be added are trained for improved inference (coming from the entire diffractive network: old and new layers). By using a patch of two layers added to an existing and fixed D^2NN design (N = 5 layers), we improved our MNIST classification accuracy to 93.39% (fig. 2.6); the state-of-the-art convolutional neural network performance has been reported as 99.60 to 99.77% [46–48].

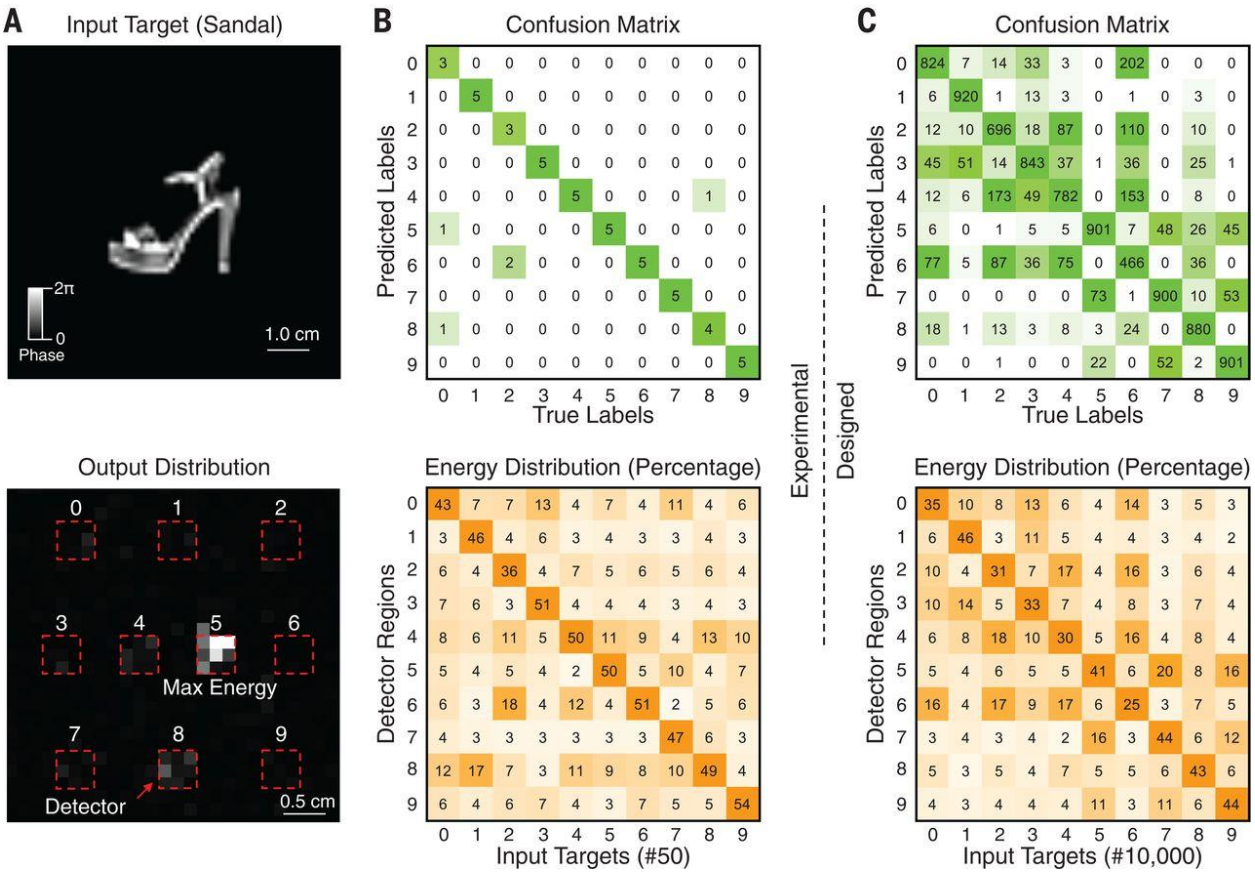

Fig. 2.4 Fashion product classifier D^{2}NN. (**A**) As an example, the output image of the 3D-printed D2NN for a sandal input (Fashion-MNIST class 5) is demonstrated. The red dashed squares represent the trained detector regions for each fashion product. Other examples of our experimental results are shown in fig. 2.14. (**B**) Confusion matrix and energy distribution percentage for our experimental results, using 50 different fashion products (five per class) that were 3D-printed, selected among the images for which numerical testing was successful. (**C**) Same as (**B**), except summarizing our numerical testing results for 10,000 different fashion products (~1000 per class), achieving a classification accuracy of 81.13% using a five-layer design. By increasing the number of diffractive layers to 10, our classification accuracy increased to 86.60% (fig. 2.9).

Following these numerical results, we 3D-printed our five-layer D^{2}NN design (Fig. 2.2A), with each layer having an area of 8 cm by 8 cm, followed by 10 detector regions defined at the output plane of the diffractive network (Figs. 2.1B and 2.3A). We then used continuous-wave

illumination at 0.4 THz to test the network's inference performance (Figs. 2.2, C and D). Phase values of each layer's neurons were physically encoded using the relative thickness of each 3D-printed neuron. Numerical testing of this five-layer D^2NN design achieved a classification accuracy of 91.75% over ~10,000 test images (Fig. 2.3C). To quantify the match between these numerical testing results and our experiments, we 3D-printed 50 handwritten digits (five different inputs per digit), selected among the same 91.75% of the test images for which numerical testing was successful. For each input object that is uniformly illuminated with the terahertz source, we imaged the output plane of the D^2NN to map the intensity distribution for each detector region that is assigned to a digit. The results (Fig. 2.3B) demonstrate the success of the 3D-printed diffractive neural network and its inference capability: The average intensity distribution at the output plane of the network for each input digit clearly reveals that the 3D-printed D^2NN was able to focus the input energy of the beam and achieve a maximum signal at the corresponding detector region assigned for that digit. Despite 3D-printing errors, possible alignment issues, and other experimental error sources in our setup, the match between the experimental and numerical testing of our five-layer D^2NN design was found to be 88% (Fig. 2.3B). This relatively small reduction in the performance of the experimental network compared to our numerical testing is especially pronounced for the digit 0 because it is challenging to 3D-print the large void region at the center of the digit. Similar printing challenges were also observed for other digits that have void regions; e.g., 6, 8, and 9 (Fig. 2.3B).

Next, we tested the classification performance of D^2NN framework with a more complicated image dataset—i.e., the Fashion-MNIST dataset [49], which includes 10 classes, each representing a fashion product (t-shirts, trousers, pullovers, dresses, coats, sandals, shirts, sneakers, bags, and

ankle boots; see fig. 2.7 for sample images). In general, for a coherently illuminated D^{2}NN, we can use the amplitude and/or phase channels of the input plane to represent data to be classified

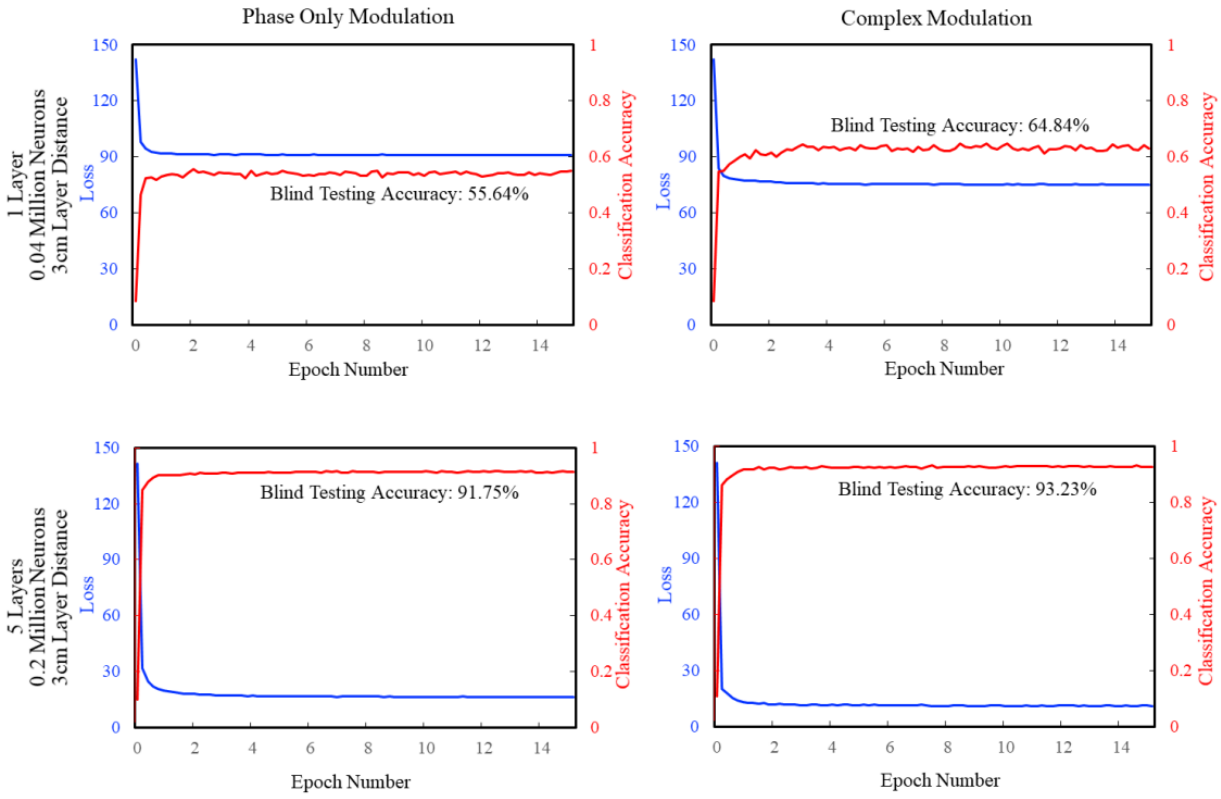

Fig. 2.5 MNIST training convergence plots A phase-only modulation D^{2}NN (left column) and a complex valued (i.e., phase and amplitude) modulation D^{2}NN (right column) as a function of the number of diffractive layers (N = 1 and 5) and the number of neurons used in the network. The y-axis values in each plot report the MNIST digit classification accuracy and the loss values as a function of the epoch number for the testing datasets. For the same number of diffractive layers, using complex-valued modulation and increasing the spacing between each layer increase the number of connections of the diffractive network, further helping to improve its inference success (also see Fig. 2.8, top two rows). For N=1, layer distance (3cm) refers to the distance between the sample/output plane and the diffractive layer. The same physical neuron size was used in each case, matching the MNIST D^{2}NN design reported in our main text. For each class, the detector width was 4.8 mm. We also obtained similar conclusions for the Fashion MNIST dataset results reported in Fig. 2.8.

or processed. In our digit classification results reported earlier, input objects were encoded by using the amplitude channel, and to demonstrate the utility of the phase channel of the network input, we encoded each input image corresponding to a fashion product as a phase-only object modulation. Our D^2NN inference results (as a function of the number of layers, neurons, and connections) for classification of fashion products are summarized in figs. 2.8 and 2.9. To provide an example of our performance, a phase-only and a complex-valued modulation D^2NN with $N = 5$ diffractive layers (sharing the same physical network dimensions as the digit classification D^2NN shown in Fig. 2.2A) reached an accuracy of 81.13 and 86.33%, respectively (fig. 2.8). By increasing the number of diffractive layers to $N = 10$ and the total number of neurons to 0.4 million, our classification accuracy increased to 86.60% (fig. 2.9). For convolutional neural net–based standard deep learning, the state-of-the-art performance for Fashion-MNIST classification accuracy has been reported as 96.7%, using ~8.9 million learnable parameters and ~2.5 million neurons[50].

To experimentally demonstrate the performance of fashion product classification using a physical D^2NN, we 3D-printed our phase-only five-layer design and 50 fashion products used as test objects (five per class) on the basis of the same procedures employed for the digit classification diffractive network (Figs. 2.2A and 2.3), except that each input object information was encoded in the phase channel. Our results are summarized in Fig. 2.4, revealing a 90% match between the experimental and numerical testing of our five-layer D^2NN design, with five errors out of 50 fashion products. Compared with digit classification (six errors out of 50 digits; Fig. 2.3), this experiment yielded a slightly better match between the experimental and numerical testing results (despite the more challenging nature of Fashion-MNIST dataset), perhaps because we used the phase channel, which does not suffer from the challenges

associated with 3D-printing of void regions [such as in digits 0, 6, 8, and 9 (Fig. 2.3)], to encode input image information for fashion products.

Next, we tested the performance of a phase-only D^2NN, composed of five 3D-printed transmission layers to implement amplitude imaging (Fig. 2.2B). This network was trained using the ImageNet database[51] to create a unit-magnification image of the input optical field amplitude at its output plane (~9 cm by 9 cm)—that is, the output image has the same physical size as the input object. As illustrated in fig. 2.10, A and C, the trained network initially connects every amplitude point at the input plane to various neurons and features of the following layers, which then focus the light back to a point at the output (i.e., image) plane, which is, as expected, quite different than the case of free-space diffraction (i.e., without the presence of the diffractive network), illustrated in fig. 2.10, B and D.

After training and blind testing, which served to numerically prove the imaging capability of the network (figs. 2.10 and 2.11), we then 3D-printed this designed D^2NN. Using the same experimental setup shown in Fig. 2.2, C and D, we imaged the output plane of the 3D-printed D^2NN for various input objects that were uniformly illuminated by continuous-wave radiation at 0.4 THz. Figure 2.12 summarizes our experimental results achieved with this 3D-printed D^2NN, which successfully projected unit-magnification images of the input patterns at the output plane of the network, learning the function of an imager, or a physical auto-encoder. To evaluate the point spread function of this D^2NN, we imaged pinholes with different diameters (1, 2, and 3 mm), which resulted in output images, each with a full width at half maximum of 1.5, 1.4, and 2.5 mm, respectively (fig. 2.12B). Our results also revealed that the printed network can resolve a linewidth of 1.8 mm at 0.4 THz (corresponding to a wavelength of 0.75 mm in air), which is

slightly worse in resolution compared with the numerical testing of our D^{2}NN design, where the network could resolve a linewidth of ~1.2 mm (fig. 2.11C). This experimental degradation in the performance of the diffractive network can be due to factors such as 3D-printing errors, potential misalignments, and absorption-related losses in the 3D-printed network.

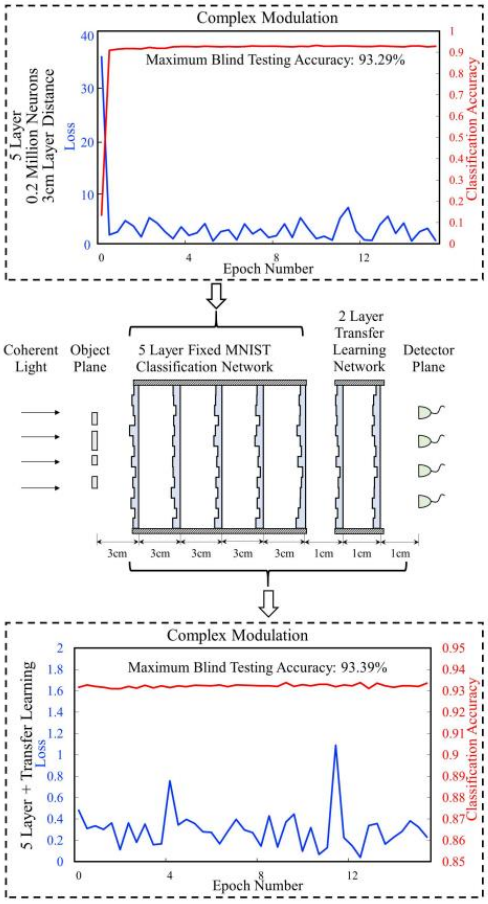

Fig. 2.6 Lego-like transfer learning approach (Top) MNIST training convergence plot of a complex-valued modulation D^{2}NN for N = 5 layers and 0.2 million neurons in total. The y-axis values report the MNIST digit classification accuracy and the loss values as a function of the epoch number for the testing dataset. (**Middle**) We illustrate a Lego-like physical transfer learning behavior for D^{2}NN framework, i.e., additional layers are patched to an existing D^{2}NN to improve its inference performance. In this example shown here, we trained 2 additional layers that were placed right at the 23 exit of an existing (i.e., fixed) 5-layer D^{2}NN. (**Bottom**) After the training of the

additional 2 layers, the inference success of the resulting "patched" diffractive neural network has reached 93.39% for MNIST testing dataset. For each class, the detector width was 0.8 mm. Also see fig. 2.20 for a comparison of detector widths.

2.3 Discussion

Optical implementation of machine learning in artificial neural networks is promising because of the parallel computing capability and power efficiency of optical systems[52–54]. Compared with previous optoelectronics-based learning approaches[52,55–57], the D^2NN framework provides a distinctive all-optical machine learning engine that efficiently operates at the speed of light using passive components and optical diffraction. An important advantage of D^2NNs is that they can be easily scaled up using various high-throughput and large-area 3D-fabrication methods (such as soft lithography and additive manufacturing), as well as wide-field optical components and detection systems, to cost-effectively reach tens to hundreds of millions of neurons and hundreds of billions of connections in a scalable and power-efficient manner. For example, integration of D^2NNs with lensfree on-chip imaging systems[15,17] could provide extreme parallelism within a cost-effective and portable platform. Such large-scale D^2NNs may be transformative for various applications, including image analysis, feature detection, and object classification, and may also enable new microscope or camera designs that can perform specific imaging tasks using D^2NNs. To achieve these new technologies, nonlinear optical materials and a monolithic D^2NN design that combines all layers of the network as part of a 3D-fabrication method would be desirable. Among other techniques, laser lithography based on two-photon polymerization[58] can provide solutions for creating such D^2NNs.

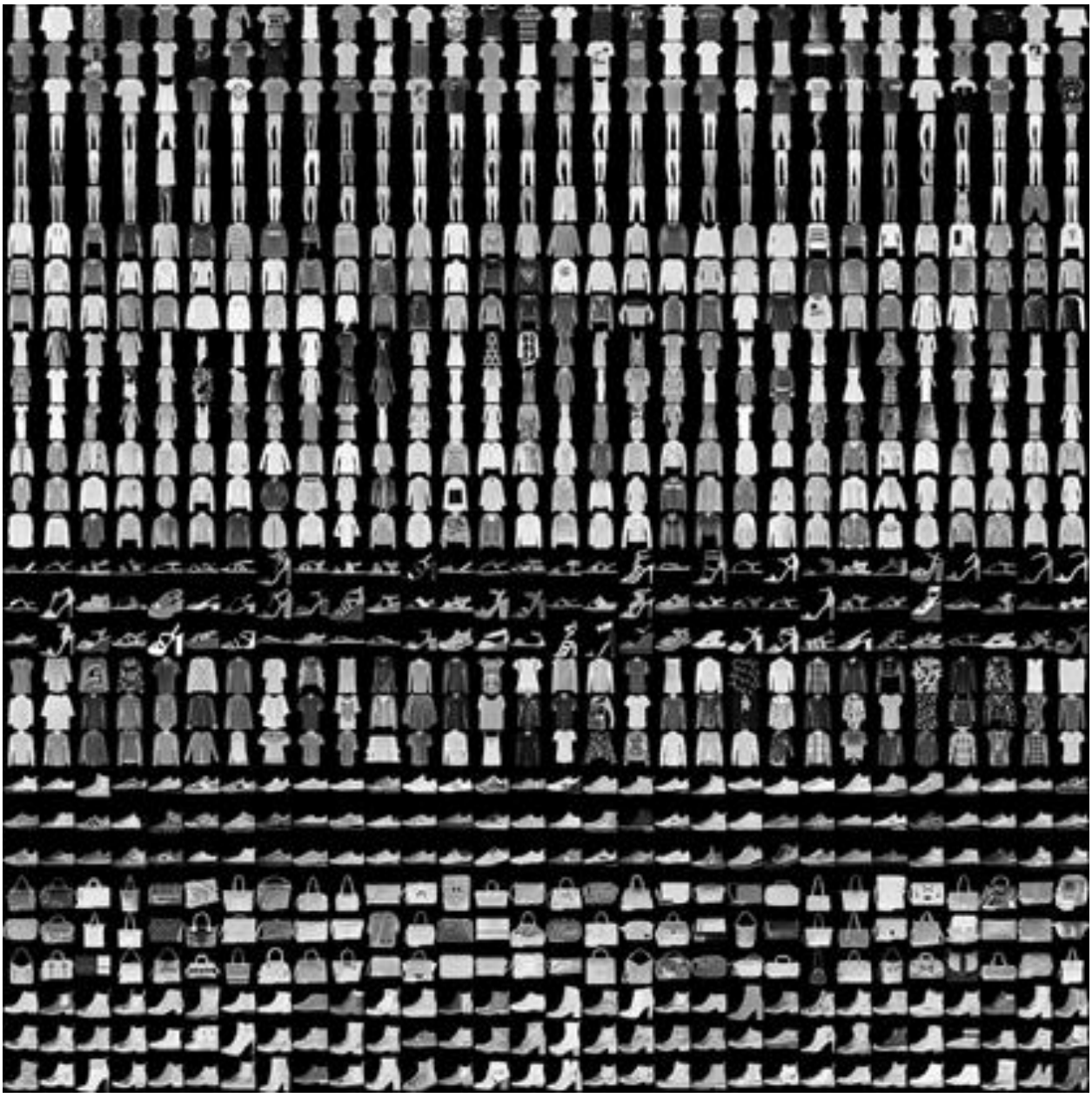

Fig. 2.7 Fashion-MNIST sample images. Some sample images for each class of the Fashion MNIST dataset.

2.4 Materials and Methods

TensorFlow-based design and 3D-printing of a D^2NN. We implemented D^2NN design using TensorFlow (Google Inc.) framework, as shown in Fig. 2.15. Because we consider coherent illumination, the input information can be encoded in the amplitude and/or phase

channels of the input plane. The free-space propagation module is implemented using the angular spectrum method. To help with the 3D-printing and fabrication of the D^2NN design, a sigmoid function was used to limit the phase value of each neuron to 0-2π. For each layer of the D^2NN, we set the neuron size to be 400 μm and 300 μm, for the digit classifier network and the lens network, respectively. With a higher resolution 3D-printer or fabrication method, smaller neurons can also be used in our D^2NN design to increase the number of neurons and connections to learn more complicated tasks. Furthermore, as illustrated in Fig. 2.11A, the number of the network layers and the axial distance between the layers are also design parameters.

At the detector/output plane, we measured the intensity of the network output, and as a loss function to train the imaging D^2NN, we used its mean square error (MSE) against the target image. The classification D^2NN was also trained using a nonlinear loss function, where we aimed to maximize the normalized signal of each target's corresponding detector region, while minimizing the total signal outside of all the detector regions (see Fig. 2.3A). We used the stochastic gradient descent algorithm, Adam[59], to back-propagate the errors and update the layers of the network to minimize the loss function. The digit classifier and lens D^2NNs were trained with MNIST[45] and ImageNet[51] datasets, respectively, and achieved the desired mapping functions between the input and output planes after 10 and 50 epochs, respectively. The training batch size was set to be 8 and 4, for the digit classifier network and the imaging network, respectively. The training phase of the fashion product classifier network shared the same details as the digit classifier network, except using the Fashion MNIST dataset[49]. The networks were implemented using Python version 3.5.0. and TensorFlow framework version 1.4.0 (Google Inc.). Using a desktop computer (GeForce GTX 1080 Ti Graphical Processing Unit, GPU and Intel(R) Core(TM) i7-7700 CPU @3.60GHz), the above-outlined TensorFlow based design of a D^2NN

architecture took ~8 and 10 hours to train for the digit classifier and the lens network, respectively.

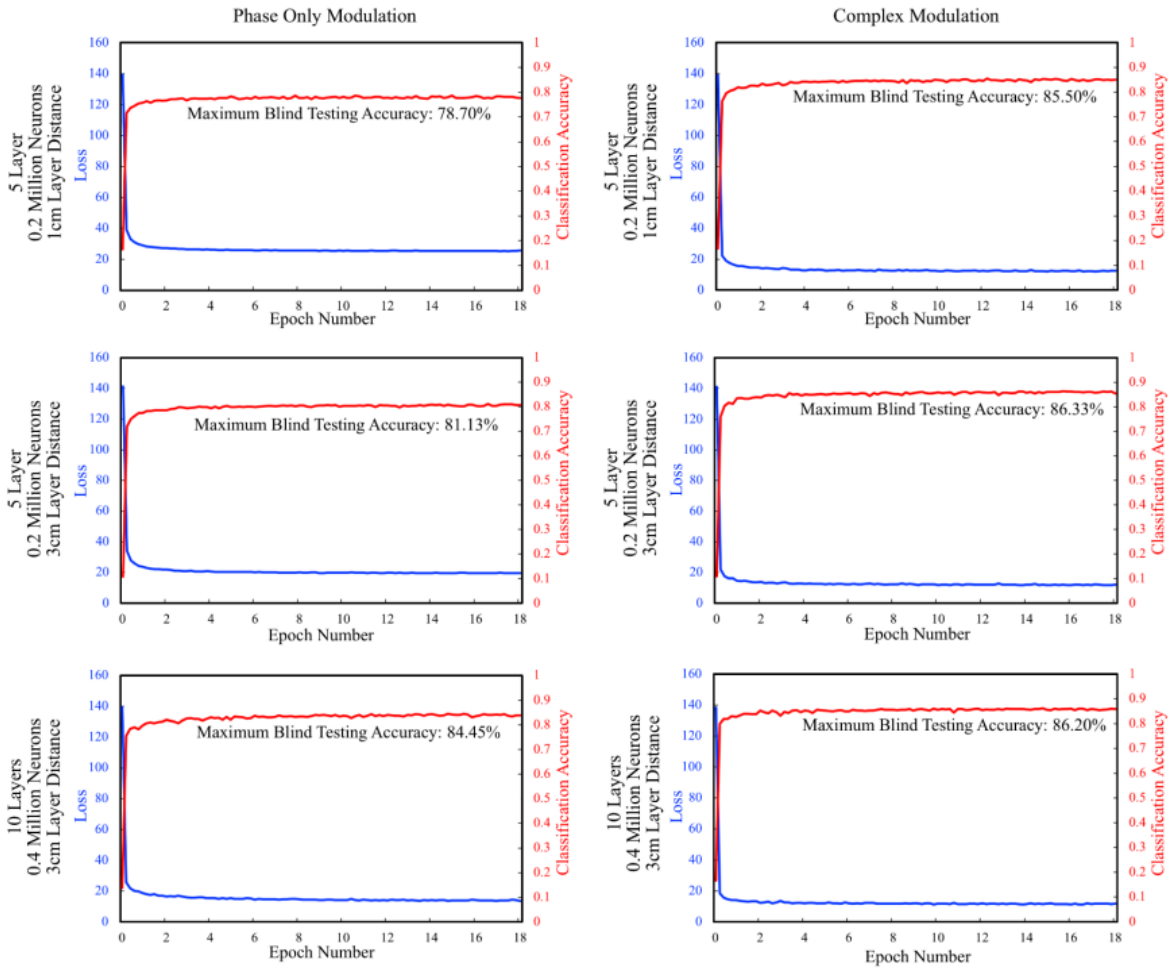

Fig. 2.8 Fashion-MNIST training convergence plots. Fashion MNIST results achieved with D^2NN framework. Training convergence plots of phase-only as well as complex-valued modulation D^2NNs (for N=5 and N=10 layers). The y-axis values in each plot report the Fashion MNIST classification accuracy and the loss values as a function of the epoch number for the testing datasets. The 1st row and 2nd row refer to the same diffractive neural network design (N=5 and 0.2 million neurons in total), except with one difference, the physical space between the layers: 1 cm vs. 3cm, respectively, which affects the number of connections in the network. As expected, the fully connected networks (with 3cm layer-tolayer distance) have better inference performance compared to the 1st row that has 1cm

layer-to-layer distance. For each class, the detector width was 4.8 mm

After the training phase of the optimized D^2NN architecture, the 3D model of the network layers to be printed was generated by Poisson surface reconstruction[60] (see Fig. 2.16). First, neurons' phase values were converted into a relative height map ($\Delta z = \lambda\phi/2\pi\Delta n$), where Δn is the refractive index difference between the 3D printing material (VeroBlackPlus RGD875) and air. The refractive index n and the extinction coefficient (k) of this 3D-printing material at 0.4 THz were measured as 1.7227 and 0.0311, respectively, which corresponds to an attenuation coefficient of $\alpha = 520.7177\, m^{-1}$. Before the 3D printing process, we also added a uniform substrate thickness of 0.5 mm to each layer of a D^2NN. A 3D mesh processing software, Meshlab[61], was used to calculate the 3D structure, which was then used as input to a 3D-printer (Objet30 Pro 3D, Stratasys Ltd, Eden Prairie, Minnesota USA). For the training of MNIST digit classifier D^2NN and Fashion-MNIST classifier D^2NN, we padded input images with zeros to fit the input aperture of the diffractive network (8 cm x 8 cm). In our THz experiments we used aluminum foil to create zero transmission regions at the input plane, to match our training settings for each D^2NN design.

Following the corresponding D^2NN design, the axial distance between two successive 3D-printed layers was set to be 3.0 cm and 4.0 mm for the digit classifier and lens networks, respectively. The larger axial distance between the successive layers of the digit classifier D^2NN increased the number of neuron connections to ~8 billion, which is approximately 100-fold larger compared to the number of the neuron connections of the imaging D^2NN, which is much more compact in depth (see Figs. 2.2(A, B)).

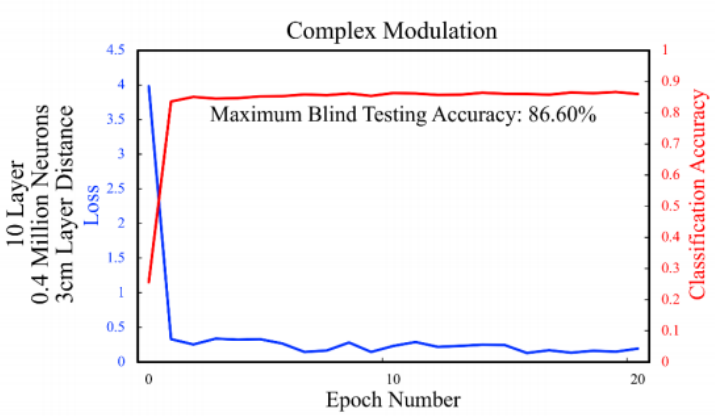

Fig. 2.9. Convergence plot of a complex-valued modulation D^{2}NN. Complex modulation D^{2}NN that has N=10 and 0.4 million neurons in total, for Fashion MNIST classification that achieves a blind testing accuracy of 86.60%. For each class, the detector width was 0.8 mm.

Terahertz Set-up. The schematic diagram of the experimental setup is given in Fig. 2.2C. The electromagnetic wave was generated through a WR2.2 modular amplifier/multiplier chain (AMC) made by Virginia Diode Inc. (VDI). A 16 dBm sinusoidal signal at 11.111 GHz (f_{RF1}) was sent as RF input signal and multiplied 36 times by AMC to generate continuous-wave (CW) radiation at 0.4 THz. We used a horn antenna compatible with WR 2.2 modular AMC. The source was electrically-modulated at 1 KHz. The illumination beam profile was characterized as a Gaussian (Fig. 2.17), and the distance between the object and the source planes was selected as 81 mm, 173 mm and 457 mm to provide a beam spot size of 20 mm, 40 mm and 104 mm, full-width half-maximum (FWHM), for the imaging D^{2}NN and the digit classification D^{2}NN, respectively. The beam passed through the input object and then the optical neural network, before reaching the output plane, which was scanned by a single-pixel detector placed on an XY positioning stage. This XY stage was built by placing two linear motorized stages (Thorlabs NRT100) vertically to allow precise control of the position of the detector. The detector scanning

step size was set to be 600 μm, 1.2 mm and 1.6 mm for the imaging lens, the digit classification D^2NN and the fashion classifier D^2NN, respectively. The distance between detector/output plane and the last layer of the optical neural network was adjusted as 3 cm and 7 mm for the classifier D^2NN and the lens D^2NN, respectively. We used a Mixer/AMC made by VDI to detect the amplitude of the transmitted wave (f_{opt}). A 10-dBm sinusoidal signal at 11.138 GHz (fRF2) was used as a local oscillator. This signal was multiplied by 36 through the multiplier and mixed with the detected signal. The mixing product ($f_{IR} = |f_{RF1} - f_{opt}|$) was obtained at 1 GHz frequency. This down-converted signal passed through an amplification stage which consisted of two 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 get rid of the noise coming from unwanted frequency bands. After this, the signal went through a low-noise power detector (Mini-Circuits ZX47-60) and 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. The dynamic range of the setup was measured as 80 dB.

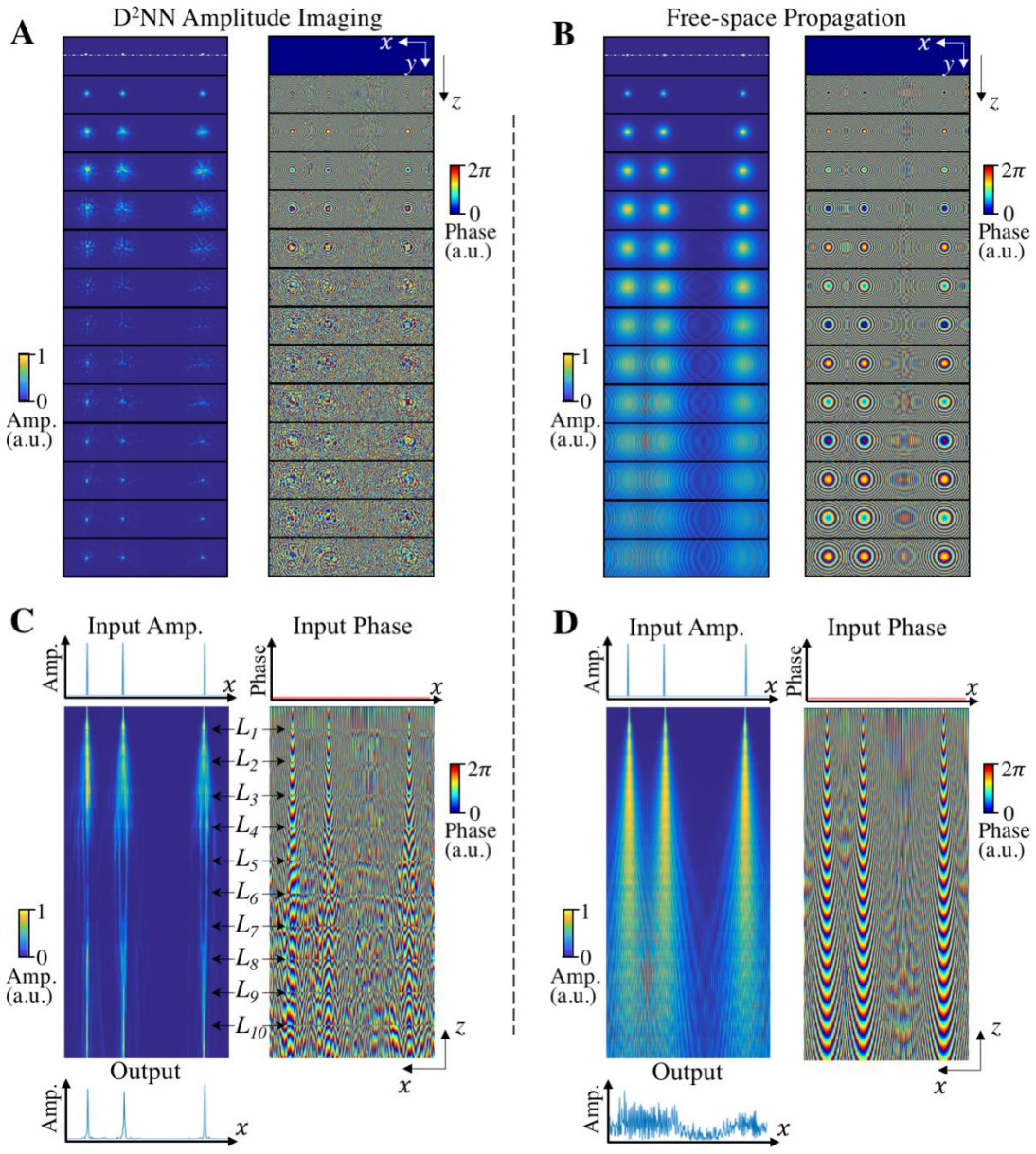

Fig. 2.10: Wave propagation within an imaging D^{2}NN. (**A, C**) To provide insights to the operation principles of a D^{2}NN, we show the amplitude and phase information of the wave that is propagating within a D^{2}NN, trained for amplitude imaging. The object was composed of 3 Dirac-delta functions spread in *x* direction. (**B, D**) Same as in (**A, C**), except without the D^{2}NN. '*L*' refers to each diffractive layer of the network. (**C**) and (**D**) show the cross-sectional view along the *z* direction indicated by the dashed lines in (**A**) and (**B**), respectively.

Wave analysis in a D^{2}NN. Following the Rayleigh-Sommerfeld diffraction equation[62], one can consider every single neuron of a given D^{2}NN layer as a secondary source of a wave that is composed of the following optical mode:

$$w_i^l(x,y,z) = \frac{z-z_i}{r^2}\left(\frac{1}{2\pi r}+\frac{1}{j\lambda}\right) exp\left(\frac{j2\pi r}{\lambda}\right), \tag{2.1}$$

where l represents the l-th layer of the network, i represents the i-th neuron located at (x_i, y_i, z_i) of layer l, λ is the illumination wavelength, $r = \sqrt{(x-x_i)^2+(y-y_i)^2+(z-z_i)^2}$ and $j = \sqrt{-1}$. The amplitude and relative phase of this secondary wave are determined by the product of the input wave to the neuron and its transmission coefficient (t), both of which are complex-valued functions. Based on this, for the l-th layer of the network, one can write the output function (n_i^l) of the i-th neuron located at (x_i, y_i, z_i) as:

$$n_i^l(x_i,y_i,z_i) = w_i^l(x_i,y_i,z_i)t_i^l(x_i,y_i,z_i)\sum_k n_k^{l-1}(x_i,y_i,z_i) = w_i^l(x_i,y_i,z_i)|A|e^{j\Delta\theta} \tag{2.2}$$

where we define $m_i^l(x_i,y_i,z_i) = \sum_k n_k^{l-1}(x_i,y_i,z_i)$ as the input wave to i-th neuron of layer l, $|A|$ refers to the relative amplitude of the secondary wave, and $\Delta\theta$ refers to the additional phase delay that the secondary wave encounters due to the input wave to the neuron and its transmission coefficient. These secondary waves diffract between the layers and interfere with each other forming a complex wave at the surface of the next layer, feeding its neurons. The transmission coefficient of a neuron is composed of amplitude and phase terms, i.e., $t_i^l(x_i,y_i,z_i) = a_i^l(x_i,y_i,z_i)exp(j\phi_i^l(x_i,y_i,z_i))$, and for a phase-only D^{2}NN architecture the amplitude $a_i^l(x_i,y_i,z_i)$ is assumed to be a constant, ideally 1, ignoring the optical losses, which is discussed in "Optical Losses in a D^{2}NN". In general, a complex-valued modulation at each network layer improves the inference performance of the diffractive network (see e.g., figs. 2.5

and 2.8).

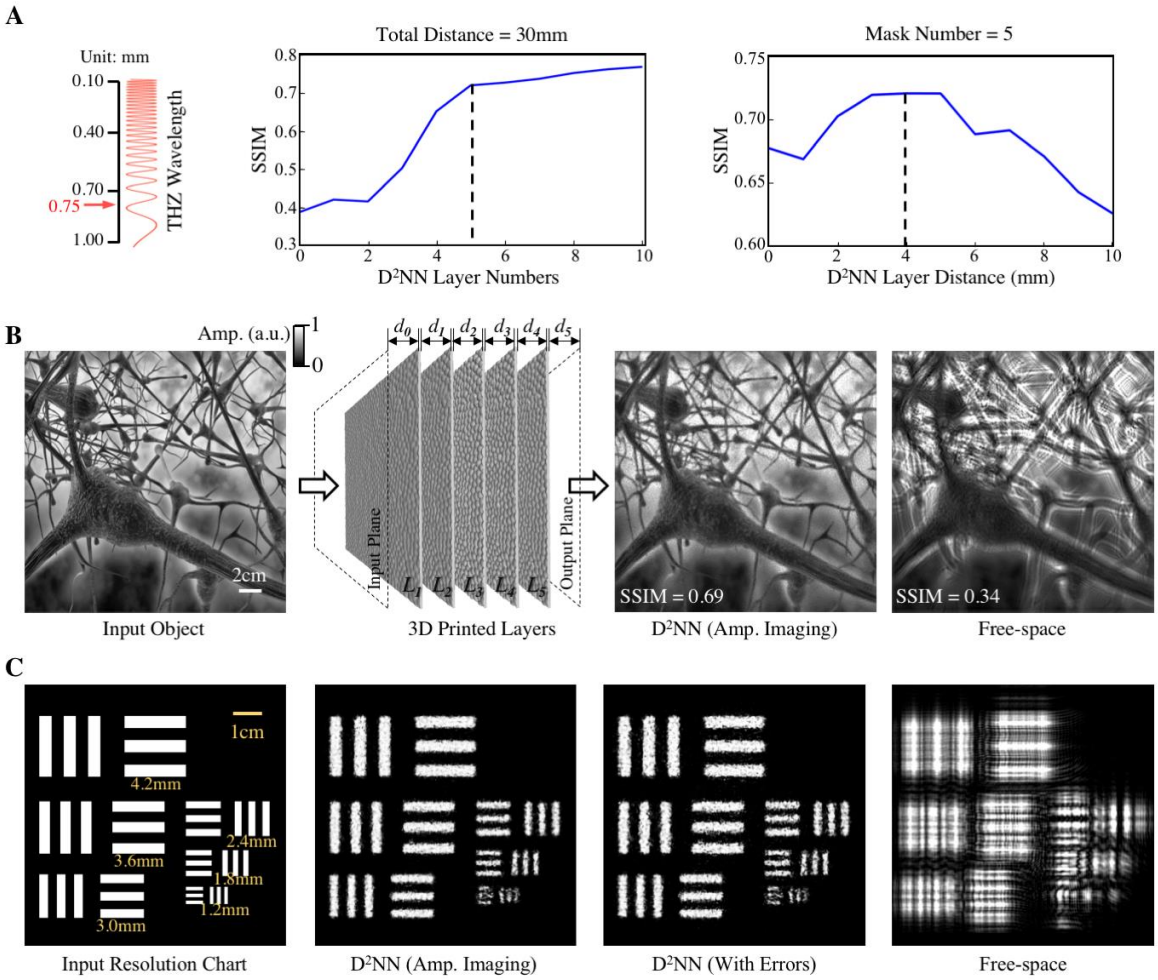

Fig. 2.11: Design of a transmissive D^{2}NN as an imaging lens. (**A**) The performance of the imaging lens D^{2}NN is optimized by tuning the physical layout of its architecture, including the number of layers (left) and the axial distance between the two consecutive layers (right). SSIM (structural similarity index) was used in this analysis, and we selected 5 layers with an axial distance of 4mm between two successive layers in order to maximize the network performance, while also minimizing its structural complexity - see the Supplementary Methods. (**B**) After the selection of the optimal neural network layout, the D^{2}NN was trained using ImageNet dataset. After its training, we blindly evaluated the performance of the resulting D^{2}NN with test images to demonstrate its success in imaging arbitrary input objects. (**C**) Blind testing results revealed that the trained D^{2}NN can resolve at its output plane a linewidth of 1.2 mm. As shown in the 3rd image on the right (D^{2}NN With Errors), the Poisson surface reconstruction

errors, absorption related losses at different layers and a random misalignment error of 0.1 mm for each layer of the network design have little effect on the imaging performance of the D^2NN. For comparison, the last image on the right shows the diffracted image at the output plane, without the presence of the D^2NN

Forward Propagation Model. The forward model of our D^2NN architecture is illustrated in Fig. 2.1A and its corresponding TensorFlow implementation is summarized in Fig. 2.15A. To simplify the notation of the forward model, we can rewrite Eq. (2.2) as follows:

$$\begin{cases} n_{i,p}^{l} = w_{i,p}^{l} \cdot t_{i}^{l} \cdot m_{i}^{l} \\ m_{i}^{l} = \sum_{k} n_{k,i}^{l-1} \\ t_{i}^{l} = a_{i}^{l} exp(j\phi_{i}^{l}) \end{cases}, \tag{2.3}$$

where *i* refers to a neuron of the *l*-th layer, and *p* refers to a neuron of the next layer, connected to neuron *i* by optical diffraction. The same expressions would also apply for a reflective D^2NN with a reflection coefficient per neuron: r_i^l. The input pattern h_k^0, which is located at layer 0 (i.e., the input plane), is in general a complex-valued quantity and can carry information in its phase and/or amplitude channels (assuming a coherent D^2NN). Without loss of generality, here we consider that the information of the input is encoded in its amplitude channel at layer 0 and the resulting wave function due to the diffraction of the illumination plane-wave interacting with the input can be written as:

$$n_{k,p}^{0} = w_{k,p}^{0} \cdot h_{k}^{0}, \tag{2.4}$$

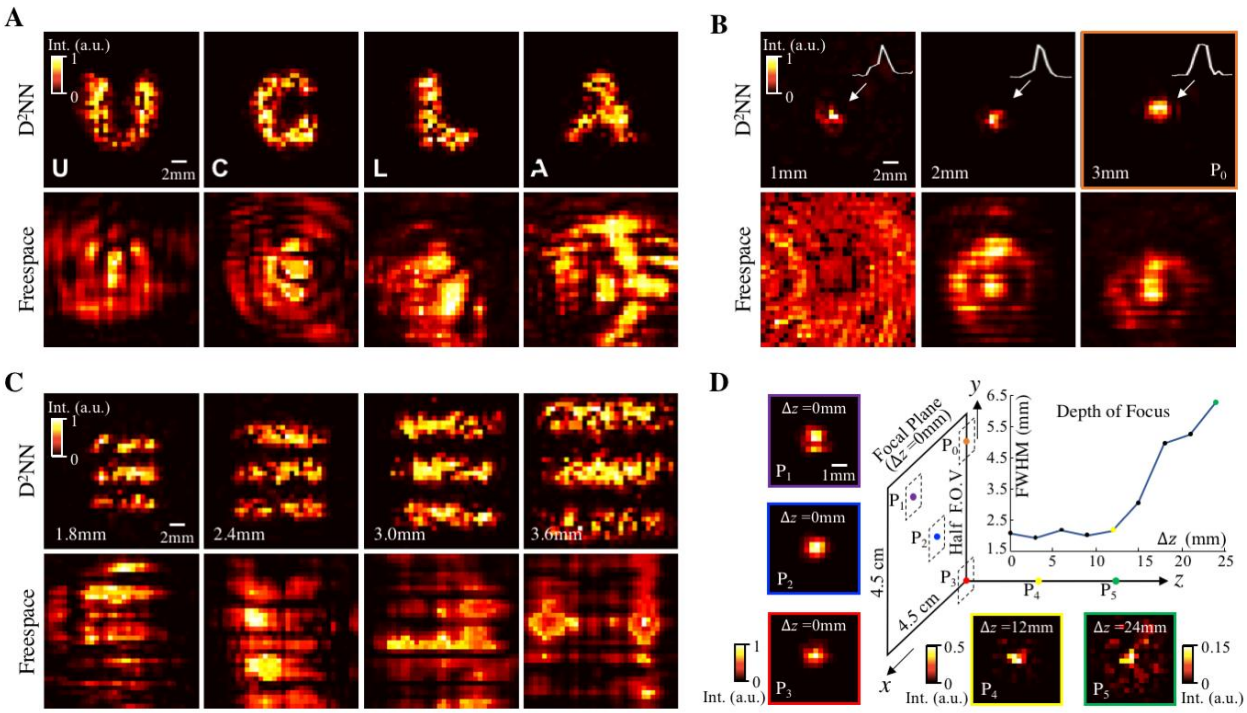

Fig. 2.12 Experimental results for imaging lens D^{2}NN. (**A**) Output images of the 3D-printed lens D^{2}NN are shown for different input objects: 'U', 'C', 'L' and 'A'. To be able to 3D-print letter 'A', the letter was slightly modified as shown in the bottom-left corner of the corresponding image panel. For comparison, free-space diffraction results corresponding to the same objects, achieved over the same sample-output plane distance (29.5 mm) without the 3D-printed network, are also shown. (**B**) Same as in (**A**), except the input objects were pinholes with diameters of 1 mm, 2 mm and 3 mm. (**C**) D^{2}NN can resolve a line-width of 1.8 mm at its output plane. (**D**) Using a 3-mm pinhole that is scanned in front of the 3D-printed network, we evaluated the tolerance of the physical D^{2}NN as a function of the axial distance. For four different locations on the input plane of the network, i.e., P_1-P_3, in (**D**) and P_0 in (**B**), we obtained very similar output images for the same 3-mm pinhole. The 3D-printed network was found to be robust to axis defocusing up to ~12 mm from the input plane. While there are various other powerful methods to design lenses[63–65], the main point of these results is the introduction of the diffractive neural network as an all-optical machine learning engine that is scalable and power-efficient to implement various functions using passive optical components, which present large degrees of freedom that can be learned through training data.

which connects the input to the neurons of layer 1. Assuming that the D^{2}NN design is composed

of *M* layers (*excluding* the input and output planes), then a detector at the output plane measures the intensity of the resulting optical field:

$$s_i^{M+1} = \left|m_i^{M+1}\right|^2. \tag{2.5}$$

The comparison of the forward model of a conventional artificial neural network and a diffractive neural network is summarized in Fig. 2.1D. Based on this forward model, the results of the network output plane are compared with the targets (for which the diffractive network is being trained for) and the resulting errors are back-propagated to iteratively update the layers of the diffractive network, which will be detailed next.

Error Backpropagation. To train a D^2NN design, we used the error back-propagation algorithm along with the stochastic gradient descent optimization method. A loss function was defined to evaluate the performance of the D^2NN output with respect to the desired target, and the algorithm iteratively optimized the diffractive neural network parameters to minimize the loss function. Without loss of generality, in our D^2NN architecture, we defined the loss function (E) using the mean square error between the output plane intensity s_i^{M+1} and the target, g_i^{M+1}:

$$E(\phi_i^l) = \frac{1}{K}\sum_k (s_k^{M+1} - g_k^{M+1})^2, \tag{2.6}$$

where *K* refers to the number of measurement points at the output plane. Based on this error definition, the optimization problem for a D^2NN design can be written as

$$\min_{\phi_i^l} E\,(\phi_i^l),\ s.t. \qquad 0 \le \phi_i^l < 2\pi \tag{2.7}$$

To apply the backpropagation algorithm for training a D^2NN, the gradient of the loss function with respect to all the trainable network variables needs to be calculated, which is then

used to update the network layers during each cycle of the training phase. The gradient of the error with respect to ϕ_i^l of a given layer *l* can be calculated as:

$$\frac{\partial E(\phi_i^l)}{\partial \phi_i^l} = \frac{4}{K}\sum_k (s_k^{M+1} - g_k^{M+1}) \cdot Real\{(m_k^{M+1})^* \cdot \frac{\partial m_k^{M+1}}{\partial \phi_i^l}\}. \quad (2.8)$$

In Eq. (2.8), $\frac{\partial m_k^{M+1}}{\partial \phi_i^l}$ quantifies the gradient of the complex-valued optical field at the output layer ($m_k^{M+1} = \sum_{k_1} n_{k_1,k}^M$) with respect to the phase values of the neurons in the previous layers, $l \leq M$. For every layer, l, this gradient can be calculated using:

$$\frac{\partial m_k^{M+1}}{\partial \phi_i^{l=M}} = j \cdot t_i^M \cdot m_i^M \cdot w_{i,k}^M, \quad (2.9)$$

$$\frac{\partial m_k^{M+1}}{\partial \phi_i^{l=M-1}} = j \cdot t_i^{M-1} \cdot m_i^{M-1} \cdot \sum_{k_1} w_{k_1,k}^M \cdot t_{k_1}^M \cdot w_{i,k_1}^{M-1}, \quad (2.10)$$

$$\frac{\partial m_k^{M+1}}{\partial \phi_i^{l=M-2}} = j \cdot t_i^{M-2} \cdot m_i^{M-2} \cdot \sum_{k_1} w_{k_1,k}^M \cdot t_{k_1}^M \cdot \sum_{k_2} w_{k_2,k_1}^{M-1} \cdot t_{k_2}^{M-1} \cdot w_{i,k_2}^{M-2}, \quad (2.11)$$

….

$$\frac{\partial m_k^{M+1}}{\partial \phi_i^{l=M-L}} = j \cdot t_i^{M-L} \cdot m_i^{M-L} \cdot \sum_{k_1} w_{k_1,k}^M \cdot t_{k_1}^M \cdots\cdots \sum_{k_L} w_{k_L,k_{L-1}}^{M-L+1} \cdot t_{k_L}^{M-L+1} \cdot w_{i,k_L}^{M-L}, \quad (2.12)$$

where, $3 \leq L \leq M$ -1. In the derivation of these partial derivatives, an important observation is that, for an arbitrary neuron at layer $l \leq M$, one can write:

$$\frac{\partial n_{k_2,k_1}^l}{\partial \phi_i^l} = \begin{cases} j \cdot t_i^l \cdot m_i^l \cdot w_{i,k_1}^l, & for\ k_2 = i \\ 0, & for\ k_2 \neq i \end{cases}, \quad (2.13)$$

where $k_{1,2}$ represent dummy variables. During each iteration of the error backpropagation, a small batch of the training data is fed into the diffractive neural network to calculate the above

gradients for each layer and accordingly update the D^2NN.

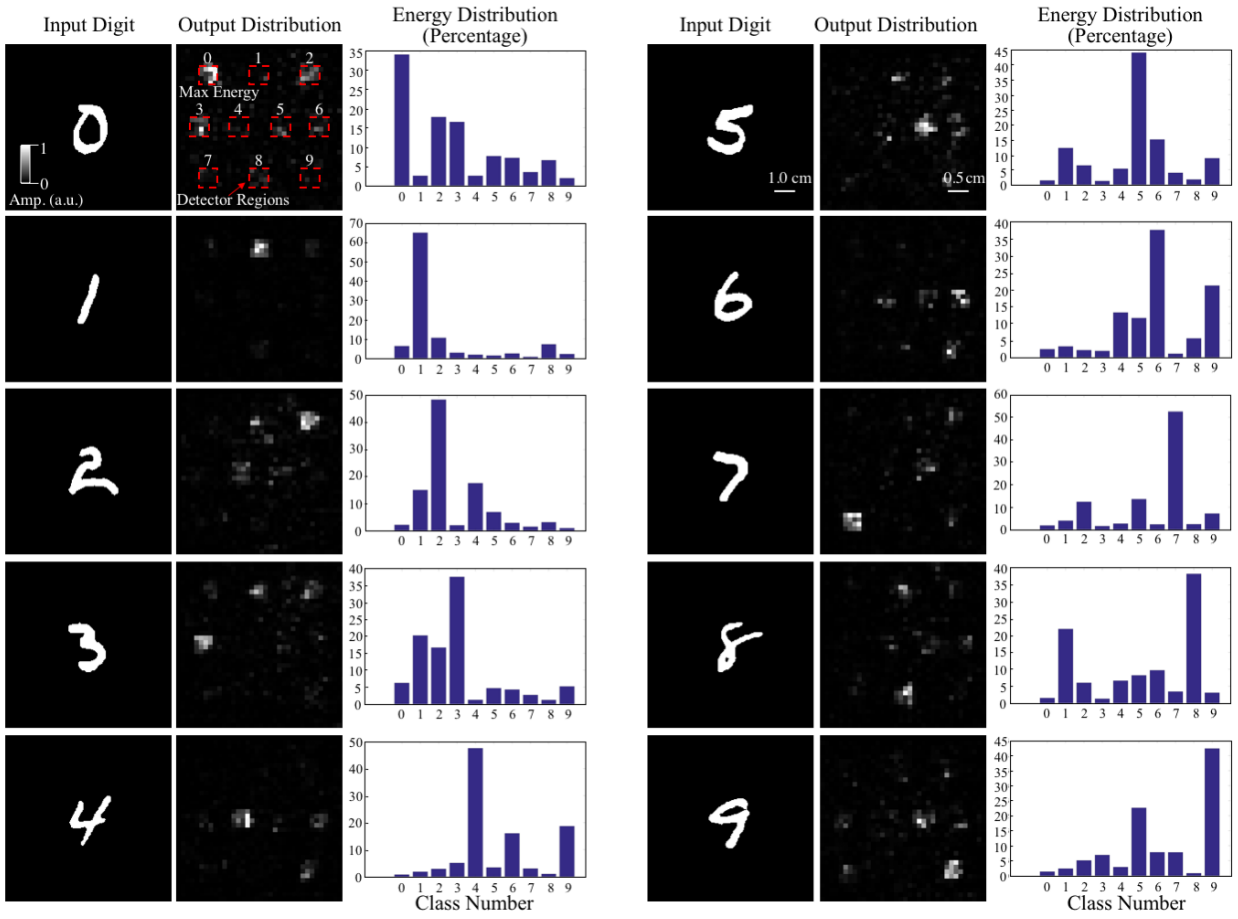

Fig. 2.13 Sample experimental results for digit classifier D^2NN. Summary of some of the experimental results achieved with our 3D-printed handwritten digit classification D^2NN. The energy distribution percentage corresponding to each digit at the output plane shows that D^2NN has the maximum energy focused on the target detector region of each digit (also see Fig. 2.3).

Comparison with standard deep neural networks. Compared to standard deep neural networks, a D^2NN is not only different in that it is a physical and all-optical deep network, but also it possesses some unique architectural differences. First, the inputs for neurons are complex-valued, determined by wave interference and a multiplicative bias, i.e., the transmission/reflection coefficient. Complex-valued deep neural networks (implemented in a

computer) with additive bias terms have been recently reported as an alternative to real-valued networks, achieving competitive results on e.g., music transcription[66]. In contrast, this work considers a coherent diffractive network modelled by physical wave propagation to connect various layers through the phase and amplitude of interfering waves, controlled with multiplicative bias terms and physical distances. Second, the individual function of a neuron is the phase and amplitude modulation of its input to output a secondary wave, unlike e.g., a sigmoid, a rectified linear unit (ReLU) or other nonlinear neuron functions used in modern deep neural networks. Although not implemented here, optical nonlinearity can also be incorporated into a diffractive neural network in various ways; see the sub-section "Optical Nonlinearity in Diffractive Neural Networks". Third, each neuron's output is coupled to the neurons of the next layer through wave propagation and coherent (or partially-coherent) interference, providing a unique form of interconnectivity within the network. For example, the way that a D^2NN adjusts its receptive field, which is a parameter used in convolutional neural networks, is quite different than the traditional neural networks, and is based on the axial spacing between different network layers, the signal-to-noise ratio (SNR) at the output layer as well as the spatial and temporal coherence properties of the illumination source. The secondary wave of each neuron will in theory diffract in all angles, affecting in principle all the neurons of the following layer. However, for a given spacing between the successive layers, the intensity of the wave from a neuron will decay below the detection noise floor after a certain propagation distance; the radius of this propagation distance at the next layer practically sets the receptive field of a diffractive neural network and can be physically adjusted by changing the spacing between the network layers, the intensity of the input optical beam, the detection SNR or the coherence length and diameter of the illumination source.

Imaging D^{2}NN Architecture. Structural similarity index, SSIM[67] values between the D^{2}NN output plane and the ground truth (i.e., target images) were calculated to optimize the

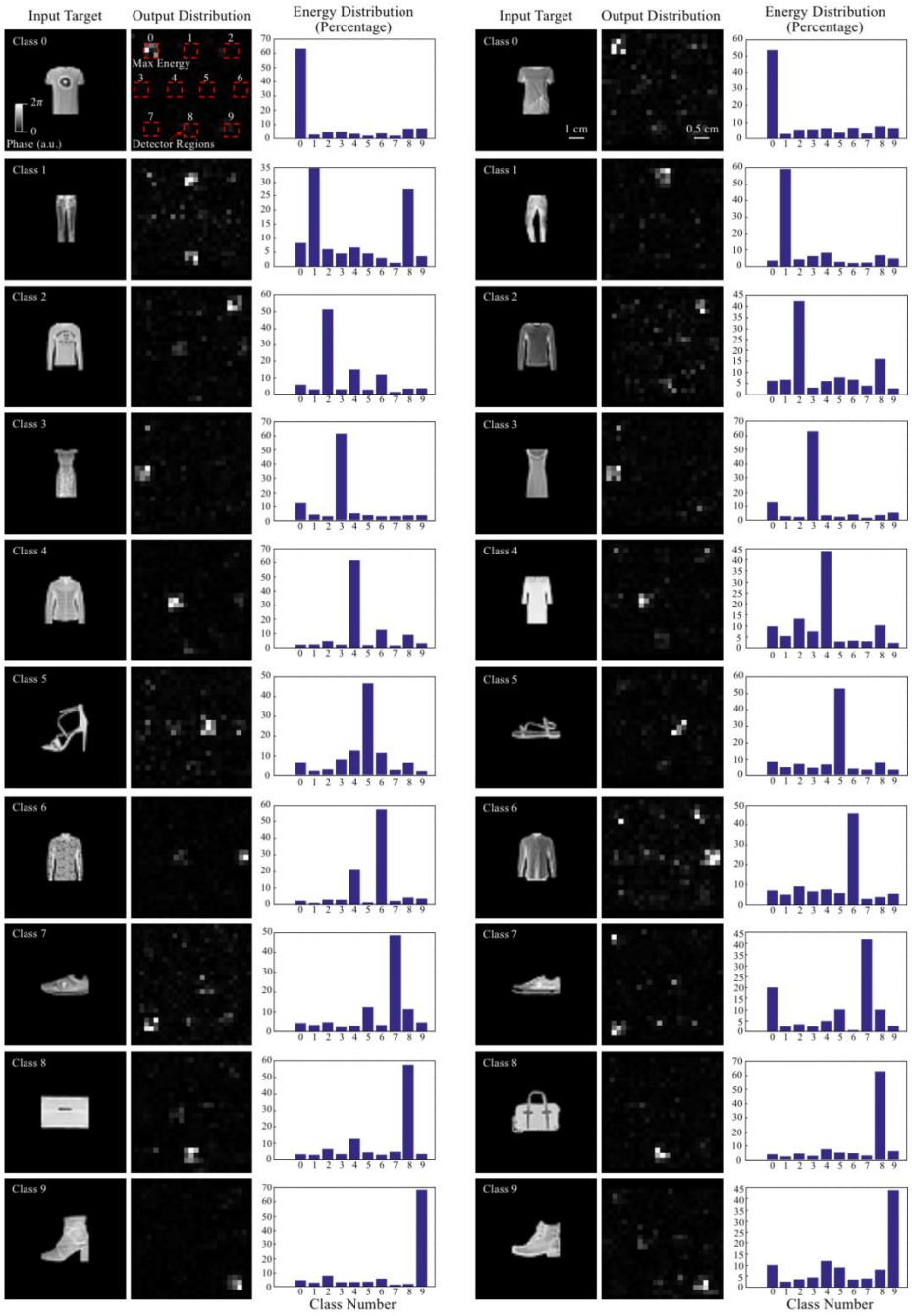

Fig. 2.14 Sample experimental results for fashion product classifier D^{2}NN. Summary of some of the experimental results achieved with our 3D-printed fashion product classification D^{2}NN. The energy distribution

percentage corresponding to each product at the output plane shows that D^2NN has the maximum energy focused on the target detector region of each product (also see Fig. 2.4)

architecture of the diffractive neural network. This way, we optimized the number of network layers and the axial distance between two consecutive layers as shown in Fig. 2.11A. The SSIM plots in Fig. 2.11A were calculated by averaging the results of 100 test images randomly selected from ImageNet.

Note also that, based on the large area of the 3D-printed imaging network layers (9×9 cm) and the short axial distance between the input (output) plane and the first (last) layer of the network, i.e., 4 mm (7 mm), one can infer that the theoretical numerical aperture of our system approaches 1 in air (see Fig. 2.2B). During the training phase, however, our diffractive network learned to utilize only part of this spatial frequency bandwidth, which should be due to the relatively large-scale of the image features that we used in the training image set (randomly selected from ImageNet database). If a higher resolution imaging system is desired, images that contain much finer spatial features can be utilized as part of the training phase to design a D^2NN that can approach the theoretical diffraction-limited numerical aperture of the system. One can also change the loss function definition used in the training phase to teach the diffractive neural network to enhance the spatial resolution; in fact deep learning provides a powerful framework to improve image resolution by engineering the loss function used to train a neural network[39,44].

Dataset Preprocessing. To train and test the D^2NN as a digit classifier, we utilized MNIST handwritten digit database[45] which is composed of 55,000 training images, 5,000 validation images and 10,000 testing images. Images were up-sampled to match the size of the D^2NN model. For the training and testing of the imaging D^2NN, we used ImageNet where we randomly

selected a subset of 2000 images. We converted each color image into grayscale and resized it to match our D^2NN design. (We should note that color image data can also be applied to D2NN framework using different approaches although we did not consider it in our work since we

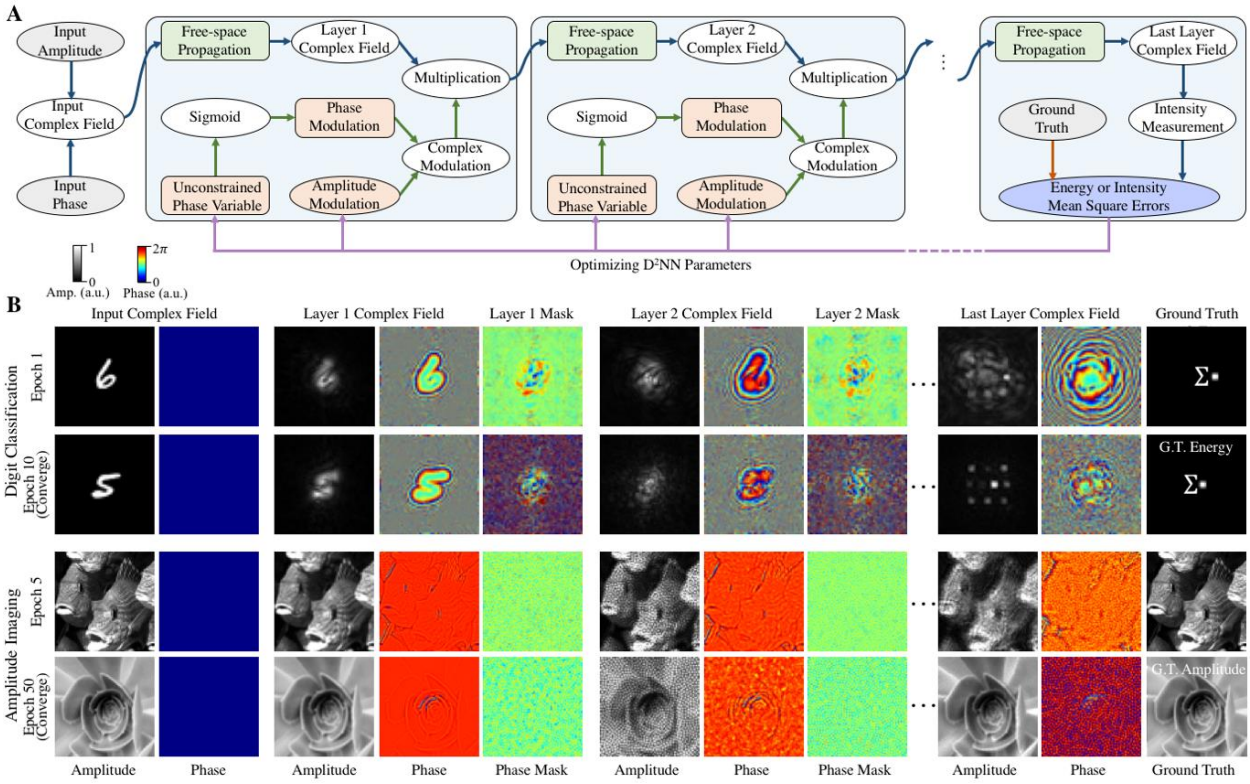

Fig. 2.15: TensorFlow implementation of a diffractive deep neural network. (**A**) The resulting complex field of free-space propagated field is multiplied with a complex modulator at each layer and is then transferred to the next layer. To help with the 3D-printing and fabrication of the D^2NN design, a sigmoid function was used to constrain the phase value of each neuron. (**B**) MNIST and ImageNet datasets were used to train the D^2NN for handwritten digit classification and imaging lens tasks, respectively. Fashion MNIST dataset was used for training the fashion product classifier D^2NN. The resulting complex fields and phase patterns of each layer are demonstrated at different epochs of the training phase.

utilized a single wavelength THz system for testing. For colorful images, as an example, Red, Green and Blue image channels can be used as parallel input channels to a diffractive neural

network) The selected images were then randomly divided into 1500 training images, 200 validation images and 300 testing images. We also obtained very similar imaging performance by using 10,000 images in the training phase; this is expected since each training image contains various spatial features at different parts of the image, all of which provide valuable patches of information for successfully training our diffractive imaging network.

To test the performance of the D^2NN digit classifier experimentally, 50 handwritten digits were extracted from MNIST test database. To solely quantify the match between our numerical testing results and experimental testing, these 3D-printed handwritten digits were selected among the same 91.75% of the test images that numerical testing was successful. The digits were up-sampled and binarized, as implemented during the training process. Binarized digits were stored as a vector image, in .svg format, before they were 3D printed. The images were then fed into Autodesk Fusion Software to generate their corresponding 3D model. To provide amplitude only image inputs to digit classifier D^2NN, the 3D-printed digits were coated with aluminum foil to block the light transmission in desired regions.

To test our D^2NN framework with a more challenging classification task, we used the Fashion MNIST database which has more complicated targets as exemplified in Fig. 2.7. Some of these target classes, such as pullovers (class 2), coats (class 4) and shirts (class 6), are very similar to each other, making it difficult for different classification methods. For example, the state-of-the-art DENSER convolutional neural network achieves 95.3% classification accuracy on Fashion MNIST dataset compared with 99.7% for MNIST dataset[49]. In order to train a D^2NN with Fashion MNIST database, we encoded the target fashion product images into the phase channel of the input plane instead of the amplitude channel. Grayscale images corresponding to

fashion products were scaled between 0 and 2π as the phase-only input to the diffractive neural network, and other details of the Fashion MNIST experiments were similar as in MNIST classification experiments.

D^{2}NN Neuron Numbers and Connectivity. D^{2}NN uses optical diffraction to connect the

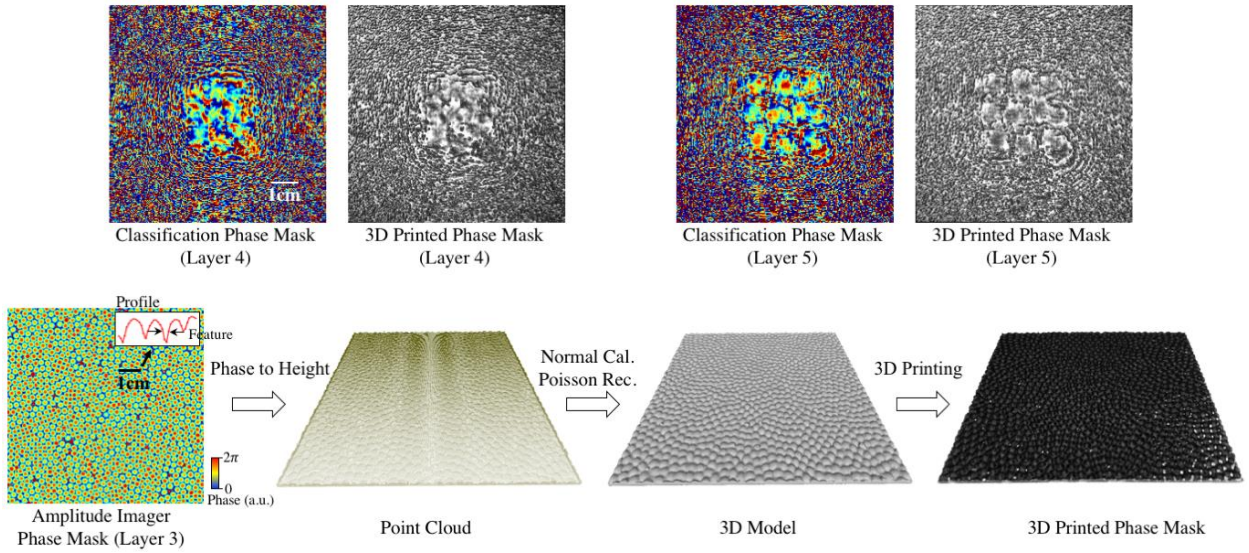

Fig. 2.16: 3D model reconstruction of a D^{2}NN layer for 3D-printing. We apply Poisson surface reconstruction to generate the 3D model of each D^{2}NN layer for 3D printing. The phase mask is first converted to a height map with the knowledge of the material refractive index, and the enclosed point cloud is formed by adding the substrate points. The 3D model is then generated by calculating the surface normal and performing the Poisson reconstruction. The final step is the 3D-printing of the D^{2}NN model.

neurons at different layers of the network. The maximum half-cone diffraction angle can be formulated as $\varphi_{max} = \sin^{-1}(\lambda f_{max})$, where $f_{max} = 1/2d_f$ is the maximum spatial frequency and d_f is the layer feature size[62]. In this work, we demonstrated the proof-of-concept of D^{2}NN architecture at 0.4 THz by using low-cost 3D-printed layers. The 3D printer that we used has a spatial resolution of 600 dpi with 0.1 mm accuracy and the wavelength of the illumination

system is 0.75 mm in air.

For the digit and fashion product classification D^2NNs, we set the pixel size to 400 μm for packing 200×200 neurons over each layer of the network, covering an area of 8 cm × 8 cm per layer. We used 5 transmissive diffraction layers with the axial distance between the successive layers set to be 3cm. These choices mean that we have a fully-connected diffractive neural network structure because of the relatively large axial distance between the two successive layers of the diffractive network. This corresponds to $200 \times 200 \times 5$=0.2 million neurons (each containing a trainable phase term) and $(200\times200)^2\times5$=8.0 billion connections (including the connections to the output layer). This large number of neurons and their connections offer a large degree-of-freedom to train the desired mapping function between the input amplitude (digit classification) or input phase (fashion product classification) and the output intensity measurement for classification of input objects.

For the imaging lens D^2NN design, the smallest feature size was ~0.9 mm with a pixel size set of 0.3 mm, which corresponds to a half-cone diffraction angle of ~25°. The axial distance between two successive layers is set to be 4 mm for 5 layers, and the width of each layer was 9 cm × 9 cm. This means the amplitude imaging D^2NN design had 300×300×5=0.45 million neurons, each having a trainable phase term. Because of the relatively small axial distance (4 mm) between the successive layers and the smaller diffraction angle due to the larger feature size, we have <0.1 billion connections in this second D^2NN design (including the connections to the output layer, which is 7 mm away from the 5^{th} layer of the diffractive network). Compared to the handwritten digit classification D^2NN, this amplitude imaging one is much more compact in the axial direction as also pictured in Fig. 2.2(A, B).

Finally, we would like to emphasize that there are some unique features of a D^2NN that make it easier to handle large scale connections (e.g., 8 billion connections as reported in Fig. 2.2A). The connectivity of a D^2NN is controlled by the size of each neuron of a given layer (defining the diffraction angle) and the axial spacing between the layers. For example, consider a 5-layer D^2NN design with a certain fixed neuron size; for this design, one can have a very low number of neural connections by closely placing the layers, one after another. On the other hand, one can also make the same design fully-connected by simply increasing the axial spacing between the layers, significantly increasing the number of connections. Interestingly, these two extreme designs (that vary significantly in their number of connections) would be identical in terms of training complexity because the computation time and complexity of digital wave propagation between layers is a not a function of the axial distance. Therefore largely spaced D^2NN layers that form a fully connected network would be identical (in terms of their computational implementation complexity) to partially-connected D^2NN designs that have shorter axial distance between the layers (also see Figure 2.8, top two rows, for an example on this comparison).

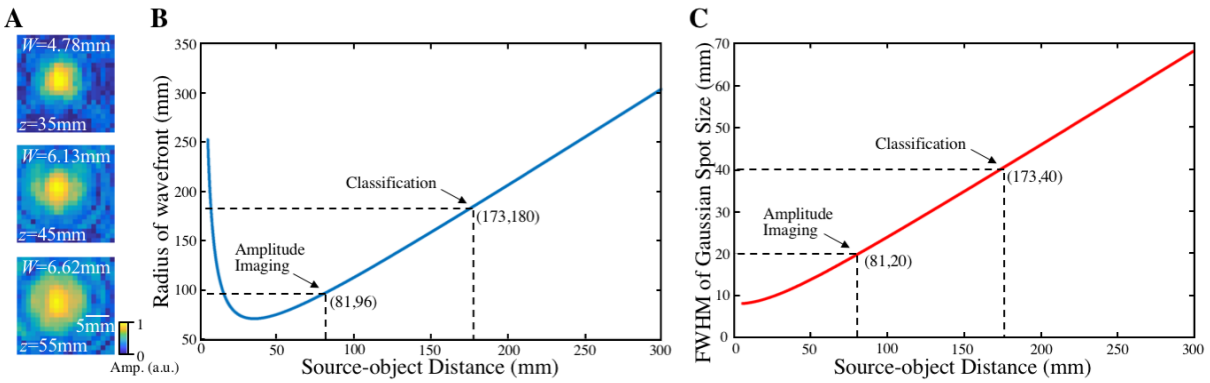

Fig. 2.17: Terahertz Source Characterization. (**A**) Beam profiles were imaged at three different axial locations to quantify the beam parameters, based on which the Terahertz light source can be approximated as a Gaussian beam. (**B, C**) The plots show the radius of the source wavefront and its full width at half maximum (FWHM) as a function

of the source-object distance. For both of the 3D-printed D^2NN designs, the illumination at the object/input plane can be approximated as a plane wave.

Performance analysis of D^2NN as a function of the number of layers and neurons. A single diffractive layer cannot achieve the same level of inference that a multi-layer D^2NN structure can perform. Multi-layer architecture of D^2NN provides a large degree-of-freedom within a physical volume to train the transfer function between its input and the output planes, which, in general, cannot be replaced by a single phase-only or complex modulation layer (employing phase and amplitude modulation at each neuron).

To expand on this, we would like to first show that, indeed, a single diffractive layer performance is quite primitive compared to a multi-layered D^2NN. As shown in Figure 2.5, a single phase-only modulation layer or even a complex modulation layer (where both phase and amplitude of each neuron are learnable parameters) cannot present enough degrees of freedom to establish the desired transfer function for classification of input images (MNIST) and achieves a much lower performance compared to a 5-layer D^2NN network, the one that we demonstrated above. In these results reported in Fig. 2.5, the same physical neuron size was used in each case, representing our 3D-printing resolution. Fig. 2.5 shows that a single layer diffractive network can only achieve 55.64% and 64.84% blind testing accuracy for phase-only and complex modulation D^2NN designs, respectively, whereas N=5 layers (with everything else being the same) can achieve 91.75% and %93.23 blind testing accuracy, respectively. The same conclusion also applies for a single layer D^2NN (N=1) that has 0.2 million neurons over the same area (assuming a higher resolution 3D-printer was available for defining smaller neurons).

Figure 2.6 further demonstrates that by using a patch of 2 layers added to an existing/fixed

D^2NN (N=5), we improved our MNIST classification accuracy to 93.39%; the state of the art convolutional neural net performance varies between 99.60%-99.77% depending on the network design[46–48]. We have obtained similar results for the Fashion MNIST dataset using N=5, 10 layers (see Figs. 2.8-2.9).

These results, summarized above, highlight that a single diffractive layer stagnates at its inference performance to modest accuracy values, and increasing the number of layers, neurons and connections of a D^2NN design provides significant improvements in its inference capability.

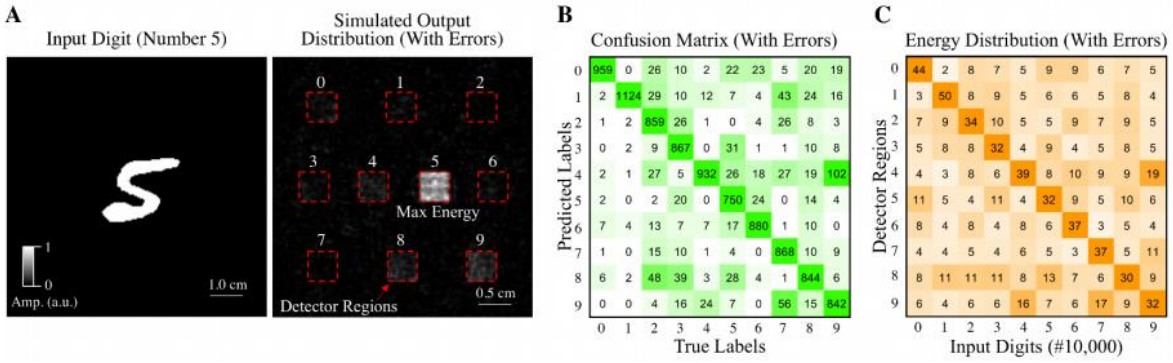

Fig. 2.18: Numerical Test Results of the Digit Classifier D^2NN Including Error Sources. (A) As an example, the output image of the digit classifier D^2NN for a handwritten input of "5" is demonstrated, where the red squares represent the trained detector regions for each digit. (**B, C**) are the same as in Fig. 3C of the main text, except they now take into account the Poisson surface reconstruction errors, absorption related losses at different layers and a random misalignment error of 0.1 mm for each layer of the network design. All these sources of error reduced the overall performance of the diffractive network's digit classification accuracy from 91.75% (Fig. 3C) to 89.25%, evaluated over 10,000 different handwritten digits (i.e., approximately 1,000 for each digit).

Error sources and mitigation strategies. There are five main potential sources of error that contribute to the performance of a 3D-printed D^2NN: (1) Poisson surface reconstruction is the first error source. After the transmission layers are trained, 3D structure of each layer is

generated through the Poisson surface reconstruction as detailed in the Methods section. However, for practical purposes, we can only use a limited number of sampling points, which distorts the 3D structure of each layer. (2) Alignment errors during the experiments form the second source of error. To minimize the alignment errors, the transmission layers and input objects are placed into single 3D printed holder. However, considering the fact that 3D printed materials have some elasticity, the thin transmission layers do not perfectly stay flat, and they will have some curvature. Alignment of THz source and detector with respect to the transmission layers also creates another error source in experiments. (3) 3D printing is the third and the most dominant source of error. This originates from the lack of precision and accuracy of the 3D printer used to generate transmission layers. It smoothens the edges and fine details on the transmission layers. (4) Absorption of each transmissive layer is another source that can deteriorate the performance of a D^2NN design. (5) The measurements of the material properties that are extensively used in our simulations such as refractive index and extinction coefficient of the 3D printed material might have some additional sources of error, contributing to a reduced experimental accuracy. It is hard to quantitatively evaluate the overall magnitude of these various sources of errors; instead we incorporated the Poisson surface reconstruction errors, absorption related losses at different layers and 0.1 mm random misalignment error for each network layer during the testing phase of the D^2NNs as shown in Figs. 2.11 and 2.18. These errors showed minor influence on the performance of the diffractive networks.

To minimize the impact of the 3D printing error, we set a relatively large pixel size, i.e. 0.4 mm and 0.3 mm for the classification and imaging D^2NNs, respectively. Furthermore, we designed a 3D-printed holder (Figs. 2.2(A, B)) to self-align the multi-layer structure of a 3D-printed D^2NN, where each network layer and the input object were inserted into their specific

slots. Based on the resolution of our 3D-printer, the misalignment error of a 3D-printed D^2NN (including its holder) is estimated to be smaller than 0.1 mm compared to the ideal positions of the neurons of a given layer, and this level of error was found to have a minor effect on the network performance as illustrated in Figs. 2.11 and 2.18.

For an inexpensive 3D-printer or fabrication method, printing/fabrication errors and imperfections, and the resulting alignment problems can be further mitigated by increasing the area of each layer and the footprint of the D^2NN. This way, the feature size at each layer can be increased, which will partially release the alignment requirements. The disadvantage of such an approach of printing larger diffractive networks, with an increased feature size, would be an increase in the physical size of the system and its input optical power requirements. Furthermore, to avoid bending of the network layers over larger areas, an increase in layer thickness and hence its stiffness would be needed, which can potentially also introduce additional optical losses, depending on the illumination wavelength and the material properties.

Optical Losses in a D^2NN. For a D^2NN, after all the parameters are trained and the physical diffractive network is fabricated or 3D-printed, the computation of the network function (i.e., inference) is implemented all-optically using a light source and optical diffraction through passive components. Therefore, the energy efficiency of a D^2NN depends on the reflection and/or transmission coefficients of the network layers. Such optical losses can be made negligible, especially for phase-only networks that employ e.g., transparent materials that are structured using e.g., optical lithography, creating D^2NN designs operating at the visible part of the spectrum. In our experiments, we used a standard 3D-printing material (VeroBlackPlus RGD875) to provide phase modulation, and each layer of the networks shown in Fig. 2.2 had on

average ~51% power attenuation at 0.4 THz for an average thickness of ~1 mm (see Fig. 2.19). This attenuation could be further decreased by using thinner substrates or by using other materials (e.g., polyethylene, polytetrafluoroethylene) that have much lower losses in THz wavelengths. One might also use the absorption properties of the neurons of a given layer as another degree of freedom in the network design to control the connectivity of the network, which can be considered as a physical analog of the dropout rate in deep network training[68]. In principle, a phase-only D^2NN can be designed by using the correct combination of low-loss materials and appropriately selected illumination wavelengths, such that the energy efficiency of the diffractive network is only limited by the Fresnel reflections that happen at the surfaces of different layers. Such reflection related losses can also be engineered to be negligible by using anti-reflection coatings on the substrates. So far, the consideration of multiple-reflections between the layers has been neglected since such waves are much weaker compared to the directly transmitted forward-propagating waves. The match between the experimental results obtained with our 3D-printed D^2NNs and their numerical simulations also supports this (see Figs. 2.3 and 2.4).

Although not considered in this manuscript since we are dealing with *passive* diffractive neural networks, diffractive networks can be created that use a physical gain (e.g., through optical or electrical pumping, or nonlinear optical phenomena, including but not limited to plasmonics and metamaterials) to explore the domain of *amplified bias terms*, i.e., $\left|t_i^l\right| > 1$ or $\left|r_i^l\right| > 1$. At the cost of additional complexity, such amplifying layers can be useful for the diffractive neural network to better handle its photon budget and can be used after a certain number of passive layers to boost up the diffracted signal, intuitively similar to e.g., optical amplifiers used in fiber optic communication links.

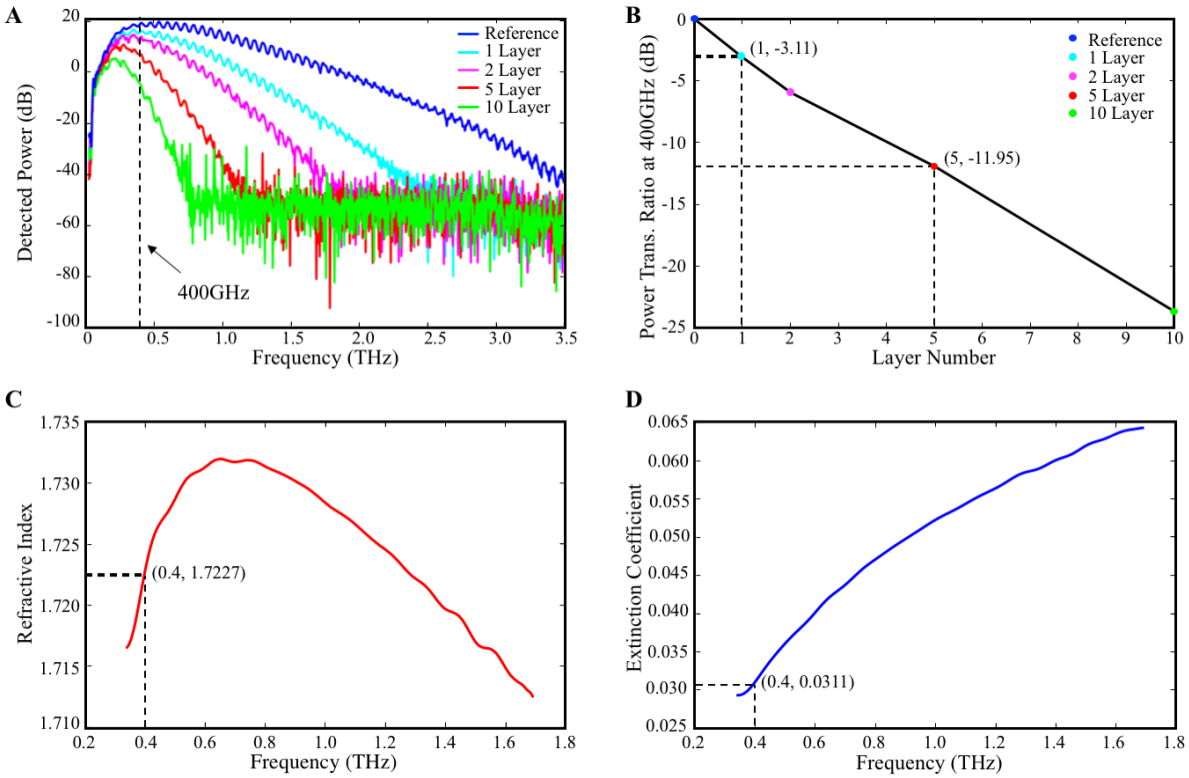

Fig. 2.19: Characterization of the 3D-printing material properties. (**A**) Our 3D-printing material (VeroBlackPlus RGD875) was characterized with a terahertz time-domain spectroscopy setup[69]. 1 mm-thick plastic layers were placed between the terahertz emitter and detector, and the transmitted field from the plastic layers was measured. The Fourier transform of the detected field was taken to calculate the detected power as a function of the frequency. The detected power levels for different numbers of 3D-printed layers are shown, revealing that the material loss increases at higher frequencies. Reference signal shows the detected power without any plastic layers on the beam path. (**B**) The power transmission ratio as a function of the number of layers is shown. The light transmission efficiency of a single 1mm-thick 3D-printed layer is $10^{-3.11/10} = 48.87\%$, and it drops to $10^{-11.95/10} = 6.38\%$ for five 1mm-thick 3D-printed layers. (**C, D**) At 0.4 THz, the refractive index and the extinction coefficient of the 3D-printing material can be calculated as 1.7227 and 0.0311, respectively. These numbers were used in the design and training of each D^2NN so that the final 3D-printed network works as designed.

Transmission and reflection modes of operation in D^2NNs. The architecture of our D2NN can be implemented in transmission or reflection modes by using multiple layers of diffractive

surfaces; in transmission (or reflection) mode of operation, the information that is transferred from one diffractive layer to the other is carried with the transmitted (or reflected) optical wave. The operation principles of D^2NN can be easily extended to amplitude-only or phase/amplitude-mixed network designs. Whether the network layers perform phase-only or amplitude-only modulation, or a combination of both, what changes from one design to another is only the nature of the multiplicative bias terms, t_i^l or r_i^l for a transmissive or reflective neuron, respectively, and each neuron of a given layer will still be connected to the neurons of the former layer through a wave-interference process, $\sum_k n_k^{l-1}(x_i, y_i, z_i)$, which provides the complex-valued input to a neuron. Compared to a phase-only D^2NN design, where $|t_i^l| = |r_i^l| = 1$, a choice of $|t_i^l| < 1$ or $|r_i^l| < 1$ would introduce additional optical losses, and would need to be taken into account for a given illumination power and detection SNR at the network output plane.

Reconfigurable D^2NN Designs. One important avenue to consider is the use of spatial light modulators (SLMs) as part of a diffractive neural network. This approach of using SLMs in D^2NNs has several advantages, at the cost of an increased complexity due to deviation from an entirely passive optical network to a reconfigurable electro-optic one. First, a D^2NN that employs one or more SLMs can be used to learn and implement various tasks because of its reconfigurable architecture. Second, this reconfigurability of the physical network can be used to mitigate alignment errors or other imperfections in the optical system of the network. Furthermore, as the optical network statistically fails, e.g., a misclassification or an error in its output is detected, it can mend itself through a transfer learning based re-training with appropriate penalties attached to some of the discovered errors of the network as it is being used. For building a D^2NN that contains SLMs, both reflection and transmission based modulator devices can be used to create an optical network that is either entirely composed of SLMs or a

hybrid one, i.e., employing some SLMs in combination with fabricated (i.e., passive) layers.

In addition to the possibility of using SLMs as part of a reconfigurable D^2NN, another option to consider is to use a given 3D-printed or fabricated D^2NN design as a fixed input block of a new diffractive network where we train only the additional layers that we plan to fabricate. Assume for example that a 5-layer D^2NN has been printed/fabricated for a certain inference task. As its prediction performance degrades or slightly changes, due to e.g., a change in the input data, etc., we can train a few new layers to be physically added/patched to the existing printed/fabricated network to improve its inference performance. In some cases, we can even peel off (i.e., discard) some of the existing layers of the printed network and assume the remaining fabricated layers as a fixed (i.e., non-learnable) input block to a new network where the new layers to be added/patched are trained for an improved inference task (coming from the entire diffractive network: old layers and new layers).

Intuitively, we can think of each D^2NN as a "Lego" piece (with several layers following each other); we can either add a new layer (or layers) on top of existing (i.e., already fabricated) ones, or peel off some layers and replace them with the new trained diffractive blocks. This provides a unique physical implementation (like blocks of Lego) for transfer learning or mending the performance of a printed/fabricated D^2NN design.

In fact, we implemented this concept of Lego design for our MNIST diffractive network and our results are summarized in Figure 2.20, demonstrating that, for example, the addition of a 6th layer (learnable) to an already trained and fixed D^2NN with N=5 improves its inference performance, coming close to the performance of a D^2NN with N=6 layers that were simultaneously trained. Also see Figure 2.6 for an implementation of the same concept for

MNIST: using a patch of 2 layers added to an existing/fixed D^2NN (N=5), we improved our MNIST classification accuracy to 93.39%. The advantage of this Lego-like transfer learning or patching approach is that already fabricated and printed D^2NN designs can be improved in performance by adding additional printed layers to them or replacing some of the existing diffractive layers with newly trained ones. This can also help us with the training process of very large network designs (e.g., $N \geq 25$) by training them in patches, making it more tractable with state of the art computers.

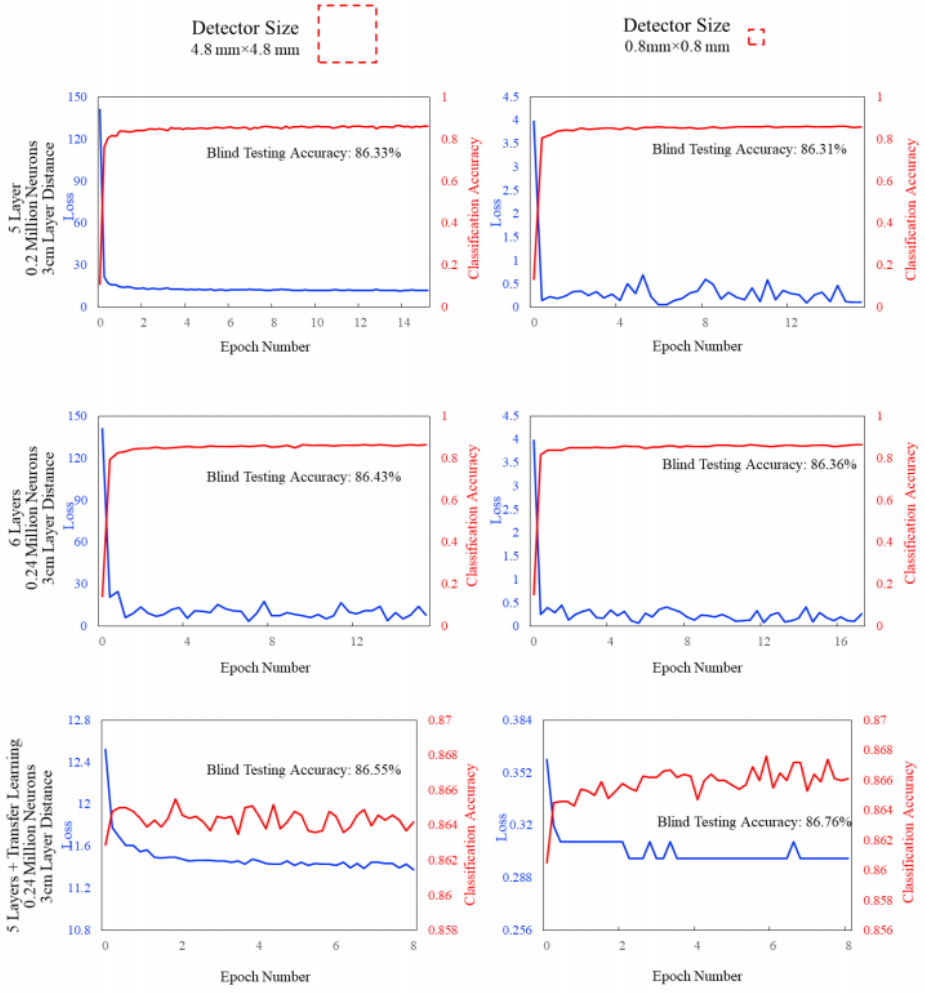

Fig. 2.20: Fashion MNIST results achieved with complex-valued D2NN framework (also see Figs. 2.8 and 2.9).

Convergence plots of D^2NNs (top and middle plots for N=5 and N=6, respectively) are shown. Bottom plots show the case for training only the 6th layer, where the first 5 layers of the network were fixed (i.e., identical to the design resulting from the top case, N=5) and the new layer was added between the 5th layer and the detector plane, 40 at equal distance from both. The layers of the N=5 and N=6 designs were separated by 3 cm from each other and the detector plane. The y-axis values in each plot report the Fashion MNIST classification accuracy and the loss values as a function of the epoch number for the training datasets. Addition of the 6th layer (learnable) to an already trained and fixed D2 NN with N=5 improves its inference performance, performing slightly better than the performance of N=6 (middle plots). Also see Fig. 2.6.

Discussion of Unique Imaging Functionalities using D^2NNs. We believe that the D^2NN framework will help imaging at the macro and micro/nano scale by enabling all-optical implementation of some unique imaging tasks. One possibility for enhancing imaging systems could be to utilize D^2NN designs to be integrated with sample holders or substrates used in microscopic imaging to enhance certain bands of spatial frequencies and create new contrast mechanisms in the acquired images. In other words, as the sample on a substrate (e.g., cells or tissue samples, etc.) diffracts light, a D^2NN can be used to project magnified images of the cells/objects onto a CMOS/CCD chip with certain spatial features highlighted or enhanced, depending on the training of the diffractive network. This could form a very compact chip-scale microscope (just a passive D^2NN placed on top of an imager chip) that implements, all-optically, task specific contrast imaging and/or object recognition or tracking within the sample. Similarly, for macro-scale imaging, face recognition, as an example, could be achieved as part of a sensor design, without the need for a high mega-pixel imager. For instance, tens to hundreds of different classes can potentially be detected using a modest (e.g., <1 Mega-pixel) imager chip placed at the output plane of a D^2NN that is built for this inference task.

For THz part of the spectrum, as another possible use example, various biomedical

applications that utilize THz imagers for looking into chemical sensing or the composition of drugs to detect e.g., counterfeit medicine, or for assessing the healing of wounds etc. could benefit from D^2NN designs to automate predictions in such THz-based analysis of specimen using a diffractive neural network.

Optical Nonlinearity in Diffractive Neural Networks. Optical nonlinearity can be incorporated into our deep optical network design using various optical non-linear materials (crystals, polymers, semiconductor materials, doped glasses, among others as detailed below). A D^2NN is based on controlling the diffraction of light through complex-valued diffractive elements to perform a desired/trained task. Augmenting nonlinear optical components is both practical and synergetic to our D^2NN framework.

Assuming that the input object, together with the D^2NN diffractive layers, create a spatially varying complex field amplitude E(x,y) at a given network layer, then the use of a nonlinear medium (e.g., optical Kerr effect based on third-order optical nonlinearity, $\chi^{(3)}$) will introduce an all-optical refractive index change which is a function of the input field's intensity, $\Delta n \propto \chi^{(3)} E^2$. This intensity dependent refractive index modulation and its impact on the phase and amplitude of the resulting waves through the diffractive network can be numerically modeled and therefore is straightforward to incorporate as part of our network training phase. Any third-order nonlinear material with a strong $\chi^{(3)}$ could be used to form our nonlinear diffractive layers: glasses (e.g., As_2S_3, metal nanoparticle doped glasses), polymers (e.g., polydiacetylenes), organic films, semiconductors (e.g., GaAs, Si, CdS), graphene, among others. There are different fabrication methods that can be employed to structure each nonlinear layer of a diffractive neural network using these materials.

In addition to 3rd order all-optical nonlinearity, another method to introduce nonlinearity into a D^2NN design is to use saturable absorbers that can be based on materials such as semiconductors, quantum-dot films, carbon nanotubes or even graphene films. There are also various fabrication methods, including standard photo-lithography, that one can employ to structure such materials as part of a D^2NN design; for example, in THz wavelengths, recent research has demonstrated inkjet printing of graphene saturable absorber[70]. Graphene-based saturable absorbers are further advantageous since they work well even at relatively low modulation intensities[71].

Another promising avenue to bring non-linear optical properties into D^2NN designs is to use nonlinear metamaterials. These materials have the potential to be integrated with diffractive networks owing to their compactness and the fact that they can be manufactured with standard fabrication processes. While a significant part of the previous work in the field has focused on second and third harmonic generation, recent studies have demonstrated very strong optical Kerr effect for different parts of the electromagnetic spectrum[72,73], which can be incorporated with our deep diffractive neural network architecture to bring all-optical nonlinearity into its operation.

Finally, one can also use the DC electro-optic effect to introduce optical nonlinearity into the layers of a D^2NN although this would deviate from all-optical operation of the device and require a DC electric-field for each layer of the diffractive neural network. This electric-field can be externally applied to each layer of a D^2NN; alternatively one can also use poled materials with very strong built-in electric fields as part of the material (e.g., poled crystals or glasses). The latter will still be all-optical in its operation, without the need for an external DC field.

Chapter 3 Terahertz Pulse Shaping Using Diffractive Networks

Parts of this chapter have previously been published in M. Veli et al. “Terahertz Pulse Shaping Using Diffractive Surfaces” Nature Communications, DOI: /10.1038/s41467-020-20268-z. In this chapter, I will introduce a pulse shaping framework that is using diffractive optical networks that is introduced in chapter 2.

Recent advances in deep learning have been providing non-intuitive solutions to various inverse problems in optics. At the intersection of machine learning and optics, diffractive networks merge wave-optics with deep learning to design task-specific elements to all-optically perform various tasks such as object classification and machine vision. Here, we present a diffractive network, which is used to shape an arbitrary broadband pulse into a desired optical waveform, forming a compact and passive pulse engineering system. We demonstrate the synthesis of various different pulses by designing diffractive layers that collectively engineer the temporal waveform of an input terahertz pulse. Our results constitute the first demonstration of direct pulse shaping in terahertz spectrum, where the amplitude and phase of the input wavelengths are independently controlled through a passive diffractive device, without the need for an external pump. Furthermore, a Lego-like physical transfer learning approach is presented to illustrate pulse-width tunability by replacing part of an existing network with newly trained diffractive layers, demonstrating its modularity. This learning-based diffractive pulse engineering framework can find broad applications in e.g., communications, ultra-fast imaging and spectroscopy.

3.1 Introduction

Inspired by neural interactions in human brain[74], artificial neural networks and deep learning have been transformative in many fields, providing solutions to a variety of data processing problems, including for example image recognition[32], natural language processing[75] and medical image analysis[33]. Data-driven training of deep neural networks has set the state-of-the-art performance for various applications in e.g. optical microscopy[33,39,76–80], holography[41,42,81–84] and sensing[85–88], among others. Beyond these applications, deep learning has also been utilized to solve inverse physical design problems arising in e.g., nanophotonics and plasmonics[89–92]. These advances cover a wide range of engineering applications and have motivated the development of new optical computing architectures[52,93–98] that aim to benefit from the low-latency, power-efficiency and parallelization capabilities of optics in the design of machine learning hardware. For example, Diffractive Deep Neural Networks (D^2NN)[99] have been introduced as an optical machine learning framework that uses deep learning methods, e.g., stochastic gradient-descent and error-backpropagation, to train a set of diffractive layers for computing a given machine learning task as the light propagates through these layers. Early studies conducted on this framework showed its statistical inference capabilities, achieving >98% numerical blind testing[100,101] accuracy for the classification of the images of handwritten digits. Recently, the D^2NN framework has also been extended to harness broadband radiation in order to design spatially-controlled wavelength de-multiplexing systems.[92]

In parallel to these recent advances at the intersection of optics and machine learning, there has been major progress in optical pulse shaping, including pulse compression for optical telecommunication[102] and pulse stretching for chirped pulse amplification[103]. Dynamic, customizable temporal waveform synthesis has been achieved using time[104–106] or frequency

domain[107–109] modulation. Among different approaches, the Fourier-transform based configuration[110], which relies on conventional optical components such as lenses to establish a mapping between the pixels of an optical modulation device and the spectral components of the input broadband light, is one of the most commonly employed techniques. In various forms of its implementation, the optical modulation device placed at the Fourier plane in between two gratings can be a dynamic component e.g., a spatial light modulator[111–114], an acousto-optic modulator[115,116], a movable mirror[117] or even a metasurface[118], offering engineered dispersion and wavefront manipulation, tailored for different applications.

However, these earlier pulse shaping techniques have restricted utility at some parts of the electromagnetic spectrum, such as the terahertz band, due to the lack of advanced optical components that can provide spatio-temporal modulation and control of complex wavefronts, covering both a broad bandwidth and a high spectral resolution at these frequencies[119,120]. As a result, direct shaping of terahertz pulses by engineering and independent control of the spectral amplitude and phase of the input wavelengths has not been achieved to date; instead, the synthesis of terahertz pulses has been generally performed indirectly through the engineering of the optical-to-terahertz converters or shaping of the optical pulses that pump terahertz sources.[121–125] Previous work also demonstrated an active device using an external pump-induced inhomogeneous medium to shape input terahertz pulses, with limited control of the spectral amplitude and phase of each wavelength due to the inherent dispersion of the medium.[126]

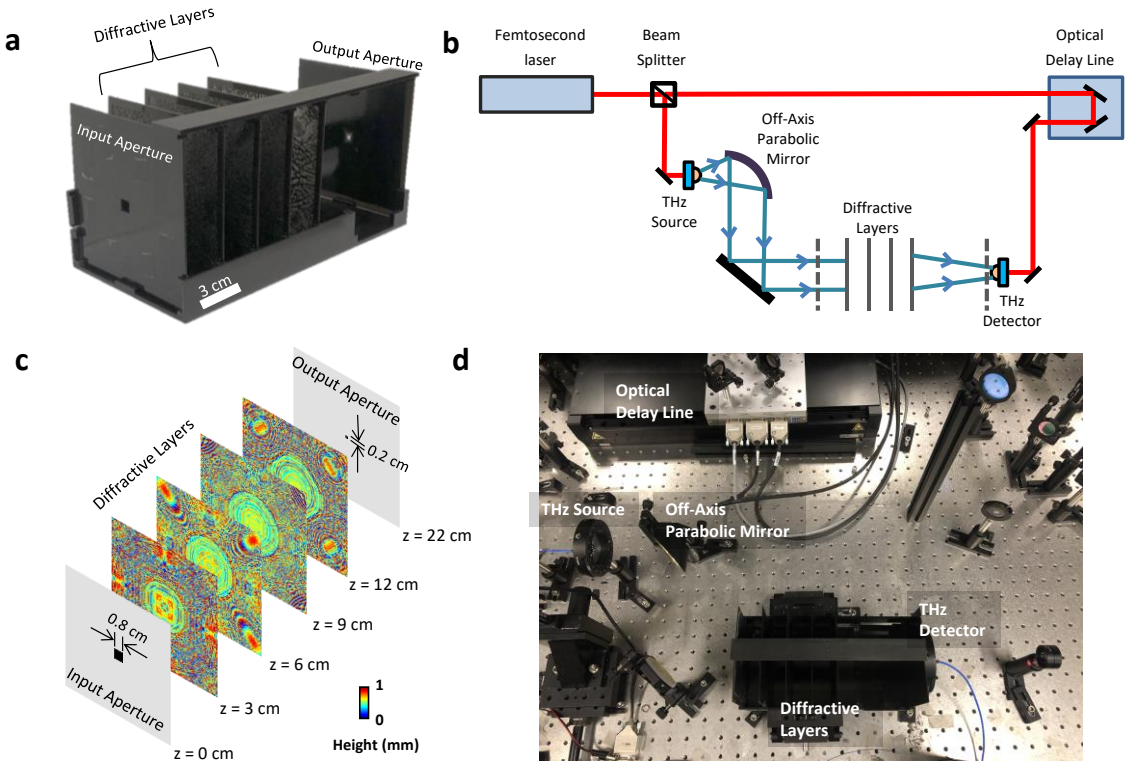

Fig. 3.1 Schematic of the pulse shaping diffractive network and a photo of the experimental setup. a 3D printed pulse shaping diffractive network that generates a square pulse with a width of 15.57 ps. **b** The schematic of the THz-TDS setup used in our experiments. The red line represents the optical path of a 780 nm femtosecond laser, and the blue line represents the terahertz beam. Dashed lines show the input and output apertures of the diffractive network. **c** The physical system layout of the pulse shaping diffractive network design. The input and output apertures are squares, with edge lengths of 0.8 cm and 0.2 cm, respectively. Gray regions on the aperture planes represent aluminum coating to block light transmission. **d** The photo of the experimental setup.

Here, we demonstrate the use of diffractive networks designed by deep learning to all-optically shape pulses by simultaneously controlling the relative phase and amplitude of each spectral component across a continuous and wide range of frequencies using only trainable diffractive layers, forming a small footprint, compact and passive pulse engineering system. This

framework uses a deep learning-based physical design strategy to devise task-specific diffractive systems that can shape various temporal waveforms of interest. Following the digital training stage in a computer, we fabricated the resulting diffractive layers (Fig. 3.1) and experimentally demonstrated the success of our pulse shaping diffractive networks by generating pulses with various temporal widths using a broadband terahertz pulse as input.

These results constitute the first demonstration of direct pulse shaping in terahertz part of the spectrum, where a complex-valued spectral modulation function that is trained using deep learning directly acts on terahertz frequencies through a passive diffractive device, without the need for an external pump. The presented learning-based approach can shape any input terahertz pulse through diffraction and is fundamentally different from previous approaches that indirectly synthesize a desired terahertz pulse through optical-to-terahertz converters or shaping of the optical pump that interacts with terahertz sources. This new capability of direct pulse shaping in terahertz band enables new opportunities that could not be explored with indirect pulse shaping approaches. For example, precise engineering and synthesis of terahertz pulses with the state-of-the-art methods is either not possible or very hard and costly to achieve, including e.g., pulsed terahertz generation through quantum cascade lasers[127–129], solid-state circuits[130,131] and particle accelerators[132]. Furthermore, the presented deep learning-based framework is quite flexible and versatile that can be used to engineer terahertz pulses regardless of their polarization state, beam shape, beam quality or aberrations of the specific terahertz generation mechanism.

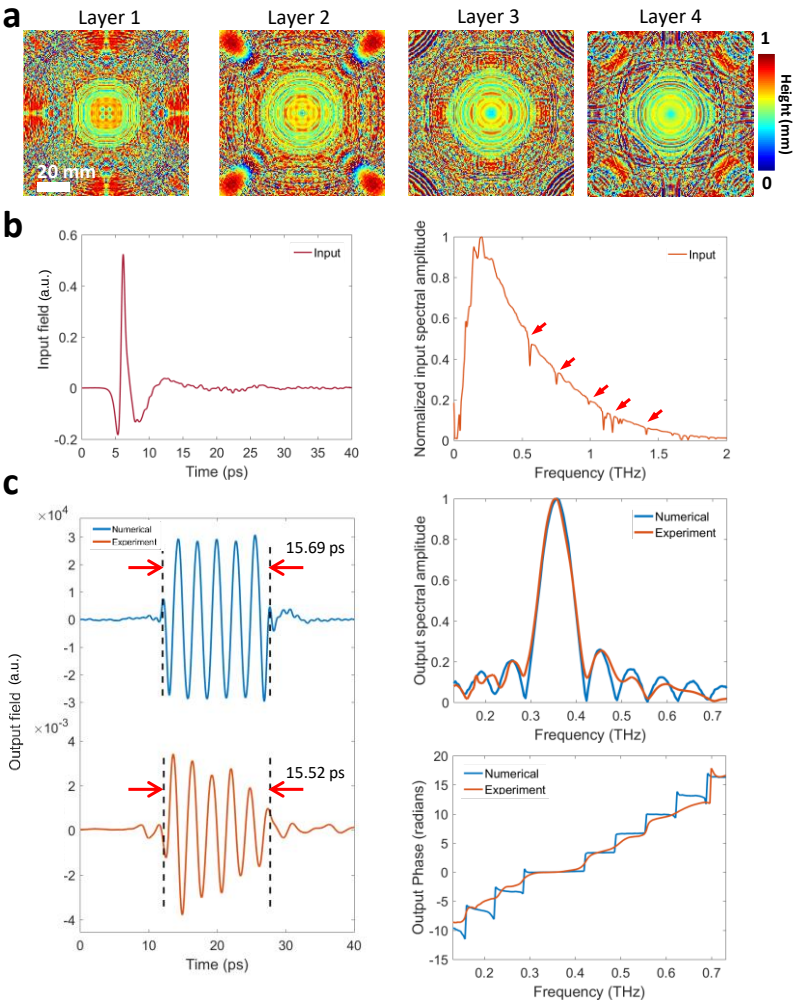

Fig. 3.2 Pulse shaping diffractive network design and output results. a The thickness profiles of the resulting diffractive layers after deep learning-based training in a computer. These diffractive layers synthesize a square pulse with a width of 15.69 ps over the output aperture for an input pulse shown in **b**. **b** Normalized input terahertz pulse measured right after the input aperture (see Fig. 1); in time-domain (left) and spectral domain (right). The red arrows on the measured spectral amplitude profile represent the water absorption bands at terahertz frequencies. **c Left**: The numerically computed (blue) and the experimentally measured (orange) output pulses in time domain. **Top right**: The normalized spectral amplitudes corresponding to the numerically computed (blue) and the experimentally measured (orange) pulses. **Bottom right**: Unwrapped spectral phase distributions computed based on the numerical forward model (blue) and the experimentally measured (orange) pulse.

Despite using passive diffractive layers, the presented pulse shaping networks offer temporal pulse-width tunability that is experimentally demonstrated by varying the inter-layer distances within a fabricated diffractive network. We also investigated a Lego-like transfer learning approach to show the modularity of the design space provided by our framework. For example, we replaced a subset of diffractive layers as part of an already trained and experimentally validated pulse shaping diffractive network design with newly trained diffractive layers to synthesize optical square pulses with different pulse-widths. All the experimental results of the 3D-fabricated pulse shaping diffractive networks presented in this work are in very good agreement with our numerically expected outputs, emphasizing the accuracy and robustness of our forward models used during the training of these diffractive networks.

In addition to engineering terahertz pulses, the fundamental design approach that is presented here can be readily adapted to different parts of the electromagnetic spectrum for shaping pulses. Finally, we believe that this study extends the engineering and precise control of electromagnetic fields through deep learning-designed diffractive networks into time-domain shaping of pulses, further motivating the development of all-optical machine learning and information processing platforms that can better harness the 4D spatio-temporal information carried by light.

3.2 Results

Synthesis of arbitrary temporal waveforms through small footprint and compact systems has been of great interest for various applications in e.g., tele-communications, ultra-fast imaging and spectroscopy, and it represents a challenging inverse design problem. Specifically, it requires accurate control of the complex-valued weights of the spectral components across a wide bandwidth and with high spectral resolution. We addressed this challenging inverse design

problem through the training of diffractive networks as shown in Fig. 3.1c. The forward training model of our diffractive networks formulates the broadband light propagation using the angular spectrum representation of optical waves[92]. Based on the complex dispersion information of a diffractive material, the *thickness* of each diffractive feature (i.e., 'neuron') of a given diffractive layer is iteratively trained and optimized through the error-backpropagation with respect to a target cost function (see the Methods section). After the convergence of this deep learning-based training in a computer, we fabricated the resulting diffractive layers (Fig. 3.1c) using a 3D-printer to physically form our pulse shaping network as shown in Fig. 3.1a. This diffractive network was then experimentally tested for its desired/targeted pulse shaping capability using a terahertz time-domain spectroscopy (THz-TDS) setup[133] that provides a noise equivalent bandwidth of 0.1-5 THz (Figs. 3.1b,d).

Each one of our pulse shaping diffractive networks consists of 4 trained layers that process the input terahertz pulse to synthesize a desired temporal waveform over an output aperture of 0.2 cm × 0.2 cm. Based on this system layout and a given input pulse profile to be shaped (Fig. 3.2b), we trained and fabricated diffractive networks that generate square pulses with different temporal widths. For example, Figure 3.2a demonstrates the diffractive layers of a pulse shaping network that was trained to generate a 15.5 ps square pulse by processing the spectrum carried by the input terahertz pulse. Figure 3.2c demonstrates the time-domain amplitude of the output waveform numerically computed (blue) based on the trained diffractive layers and the corresponding experimentally measured temporal waveform (orange), along with the associated spectral amplitude and phase distributions. The carrier frequency of the desired temporal waveform at the output was a non-learnable, predetermined parameter set to be 0.35 THz to avoid water absorption bands in the terahertz regime (depicted by the red arrows in Fig. 3.2b).

The numerically predicted output waveform (blue) in Fig. 3.2c indicates that a 4-layer diffractive network can synthesize a square temporal waveform with a pulse width of 15.69 ps without using any conventional optical components, in a compact architecture that spans approximately 250-times the carrier wavelength in the axial direction. The pulse width of the temporal waveform created by the 3D printed diffractive layers at the output aperture is measured as 15.52 ps, closely matching the numerically predicted result (15.69 ps). Similarly, a comparison of the output spectral amplitude profiles for the numerical and experimental results shows a good agreement in terms of the peak locations of the main and side lobes as well as the relative amplitude carried by each spectral component. On the other hand, an examination of the unwrapped phase profiles (experimental vs. numerical) reveals that the 3D-fabricated, physical diffractive network could not exactly create the sharp phase transitions at the expected spectral locations, but rather generated smoothened transitions. This smoothening contributes to some of the differences observed between the experimentally measured and the numerically calculated time-domain waveforms (Fig. 3.2c). The power efficiency of this diffractive network was experimentally measured as ~0.51% at the carrier frequency (f_0 = 355 GHz), quantified at the output aperture, when normalized with respect to the input; here we should emphasize that >70% of the input optical power at the carrier frequency is in fact lost due to absorption within the 3D printed diffractive layers. Therefore, to create our diffractive layers, the selection of a different fabrication material with a much lower loss (e.g., polymers such as poly-methylpentene, TPX)[134–136] can significantly boost the overall efficiency of these diffractive pulse shaping networks. Other strategies to improve our power efficiency include increasing the output aperture size and introducing additional power-related penalty terms during the training phase of the diffractive network (see Table 3.1 and Discussion section).

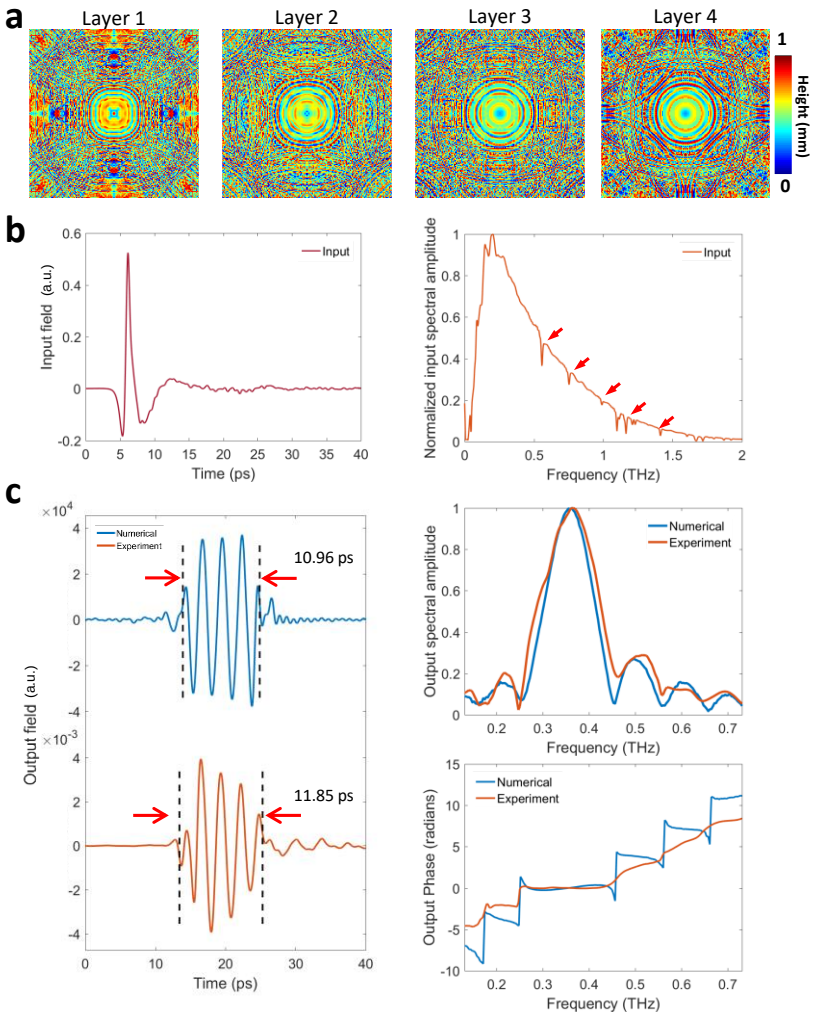

Fig. 3.3 Pulse shaping diffractive network design and output results a The thickness profiles of the resulting diffractive layers after deep learning-based training in a computer. These diffractive layers synthesize a square pulse with a width of 10.96 ps over the output aperture for an input pulse shown in **b**. **b** Normalized amplitude of the input terahertz pulse measured right after the input aperture (see Fig. 1 of the main text); in time-domain (left) and spectral domain (right). The red arrows on the measured spectral amplitude profile represent the water absorption bands at terahertz frequencies. **c Left**: The numerically computed (blue) and the experimentally measured (orange) output pulses in time domain. **Top right**: The normalized spectral amplitudes corresponding to the numerically computed (blue) and the experimentally measured (orange) pulses. **Bottom right**: Unwrapped spectral phase distributions computed based on the numerical forward model (blue) and the experimentally measured (orange) pulse.

Figure 3.3 further illustrates another diffractive network that was designed to create a narrower square pulse at its output aperture. At the end of its deep learning-based training, the numerical forward model converged to the thickness profiles shown in Fig. 3.3a in order to synthesize a 10.96 ps square pulse (blue) illustrated in Fig. 3.3c. When the diffractive layers depicted in Fig. 3.3a were 3D printed and experimentally tested using the setup shown in Fig. 3.1d, the output pulse waveform was measured to have a temporal width of 11.85 ps (orange curve in Fig 3.3c), providing a good match to our numerical results, similar to the conclusions reported in Fig. 3.2.

Beyond fabrication artefacts and misalignments observed in the 3D-printed diffractive networks, the variation of the input terahertz pulse from experiment to experiment is one of the significant contributors for any mismatch between the numerical and experimental output waveforms. The deep learning-based design of the diffractive networks shown in Figs. 3.2 and 3.3 relies on a known input terahertz pulse profile that is experimentally measured over the input aperture. To be able to take into account uncontrolled variations of the input pulse profile from run to run, we used 5 different experimentally measured input pulse profiles (dashed curves in Figs. 3.4a-b) during the training phase of each diffractive network. In the experimental testing phase, however, the terahertz input pulse (light blue curve in Figs. 3.4a-b) slightly deviated from these input pulse profiles used in the training, causing some distortions in the experimental results shown in Figs. 3.2 and 3.3, compared to their numerically computed counterparts for the same diffractive network models (also see Fig. 3.5 and Fig. 3.6).

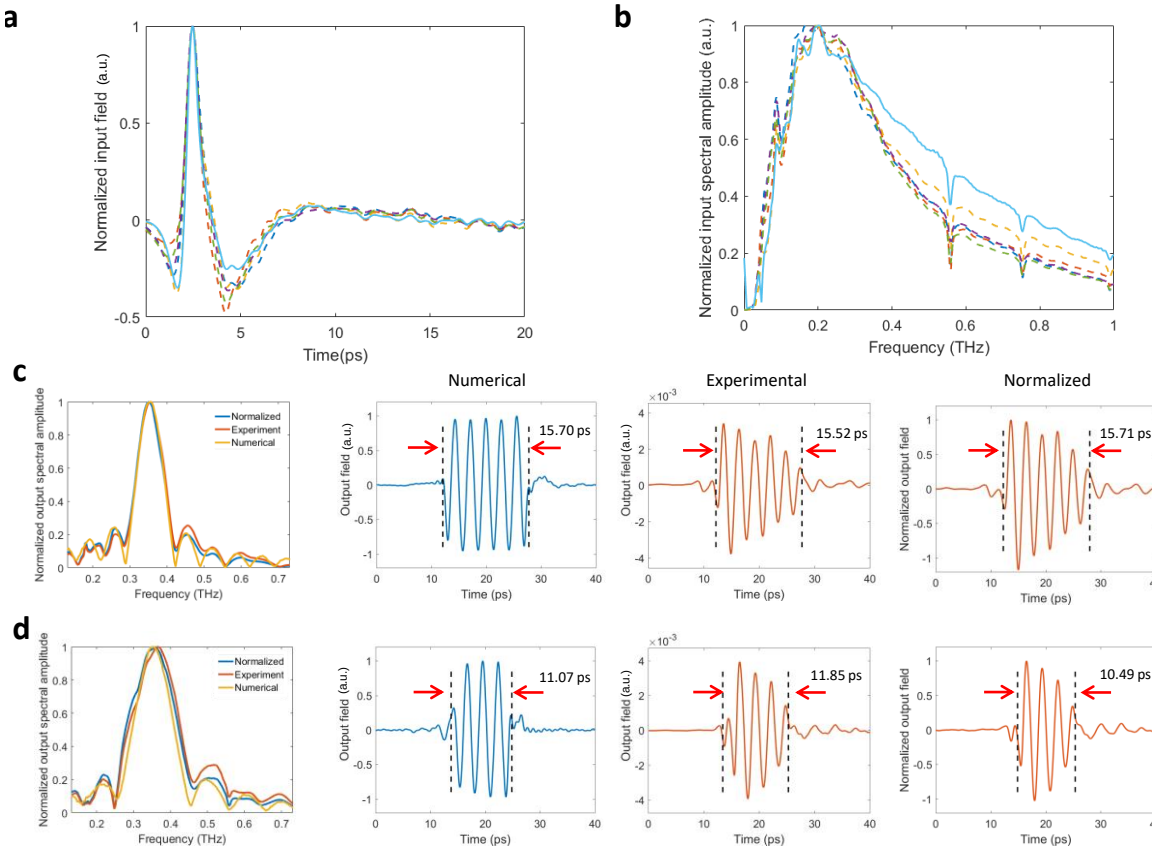

Fig. 3.4 Spectral normalization of the output pulse. a Input terahertz pulses impinging upon the diffractive network. Dashed lines represent the input pulses that were used in the training phase and the solid line represents the actual experimental input pulse used in the testing phase. **b** Normalized spectral amplitudes of the input terahertz pulses shown in **a**. **c, d** Left: normalized spectral amplitude of the output pulse obtained at the end of the training phase (yellow), compared with the experimental spectral amplitude before (orange) and after the spectral normalization step (blue). Right: numerical training output field (blue), experimental output field (orange) and the normalized output field (orange) corresponding to the desired (ground truth) square pulses with pulse-widths of **(c)** 15.49 ps and **(d)** 10.52 ps.

To shed more light onto this, next we normalized the experimentally measured spectral amplitude profiles depicted in Figs. 3.2c and 3.3c, based on the ratio between the average spectral amplitudes carried by the input pulses used in the training phase and the input pulse

measured at the experimental testing phase. This simple spectral normalization procedure nullifies the effect of input terahertz source variations from experiment to experiment and provides us an opportunity to better evaluate the accuracy of the complex-valued spectral filtering operation performed by the 3D-fabricated diffractive network. Figures 3.4c and 3.4d demonstrate the experimental spectral amplitudes and the corresponding temporal waveforms at the network output before *and* after this spectral normalization step for the diffractive networks shown in Figs. 3.2a and 3.3a, respectively. Following the spectral normalization, the width of the square pulse created by the diffractive network in Fig. 3.3a, for example, decreased from 11.85 ps to 10.49 ps, providing a better match to the 11.07 ps that is predicted by our numerical forward model (Fig. 3.4d). A similar improvement using spectral normalization was also observed for the diffractive network shown in Fig. 3.2a, almost perfectly matching its numerical counterpart in terms of the square pulse width, achieving 15.71 ps after the normalization step (Fig. 3.4c).

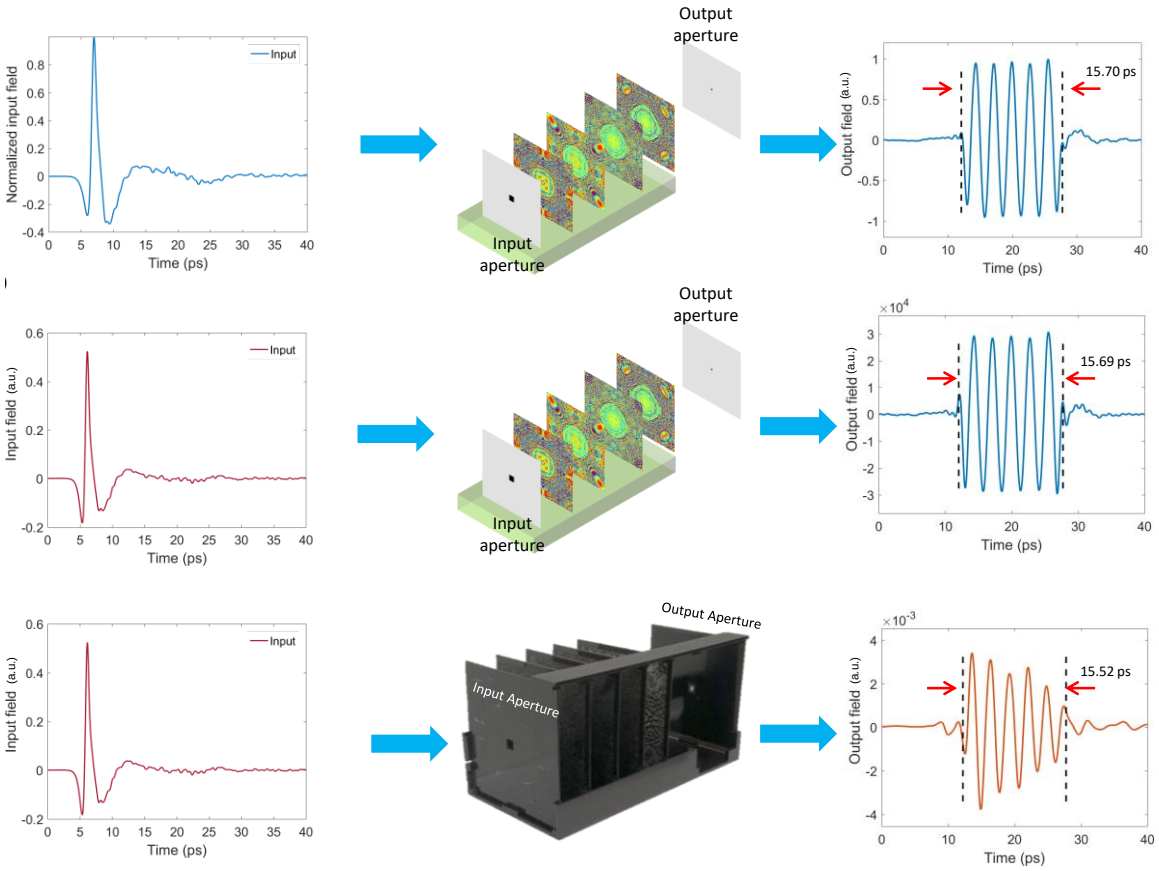

Fig. 3.5 The temporal profiles of the output pulses, 15.7 ps, resulting from the designed numerical model (a, b) and the 3D-fabricated diffractive network (c) that was trained to synthesize a 15.7 ps square pulse. (a) the input pulse is one of the pulses that were used in training phase; (b,c) the input pulse is experimentally measured at the input aperture.

These results highlight that experiment-to-experiment variability of our input terahertz pulse profile causes it to deviate from the input pulse profiles used in the training phase of our diffractive network, creating some uncontrolled errors in the output pulse profile, which can be improved significantly after the spectral normalization step, as discussed above. To further explore the pulse shaping capabilities of diffractive networks, next we trained a set of "*generic*" diffractive networks that used/assumed a *flat* input spectrum during their training in order to

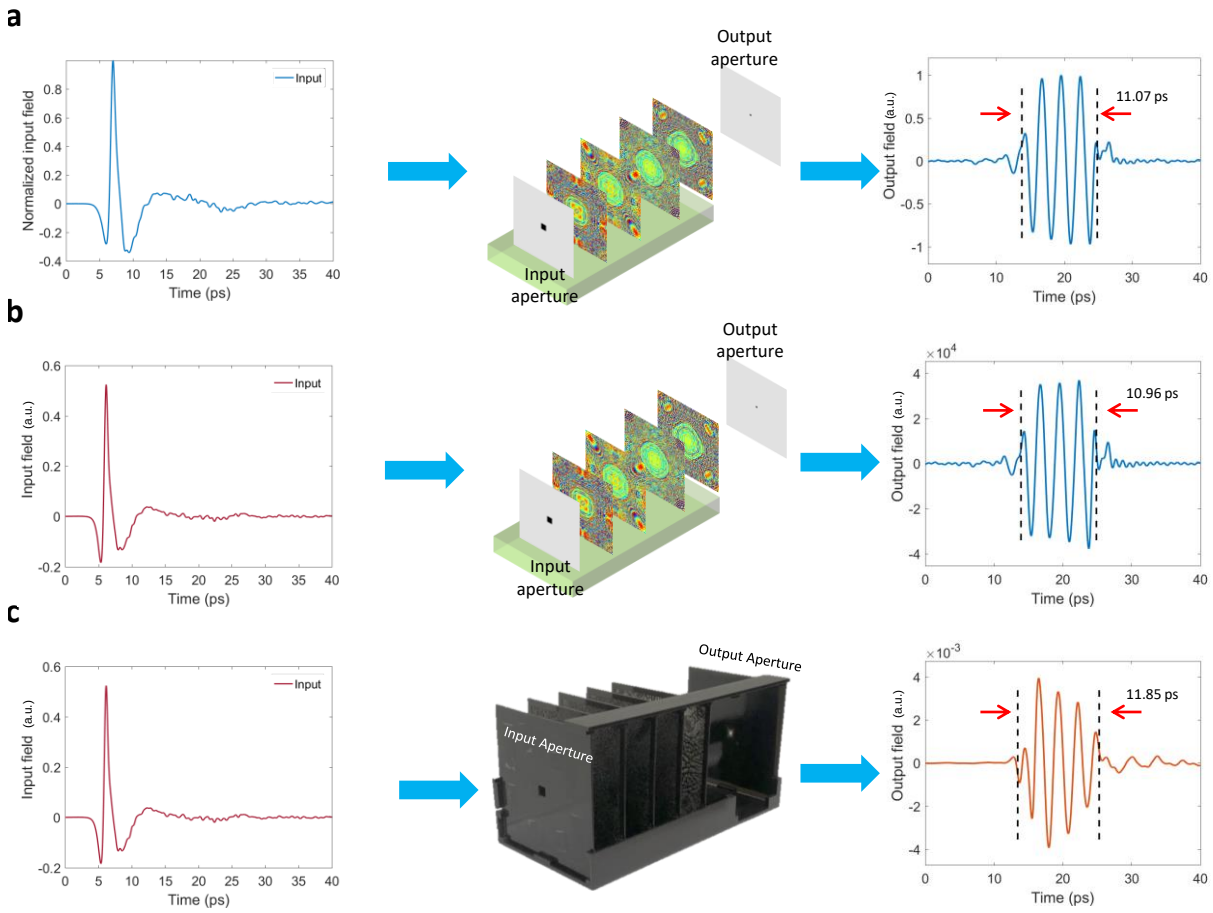

Fig. 3.6 The temporal profiles of the output pulses, 11.07 ps, resulting from the designed numerical model (a, b) and the 3D-fabricated diffractive network (c) that was trained to synthesize a 11.07 ps square pulse. (a) the input pulse is one of the pulses that were used in training phase; (b,c) the input pulse is experimentally measured at the input aperture.

achieve a desired output waveform; stated differently, a generic diffractive network is trained using an input pulse where all the wavelengths have the same spectral amplitude and phase. To accurately demonstrate the pulse shaping behavior of these generic diffractive designs that were trained with flat spectra, we used spectral normalization based on the input pulse profile, experimentally measured at each run. For example, Figures 3.7a and 3.8 show the diffractive layers of a *generic* pulse shaping network model that was trained to create a 15.5 ps square pulse.

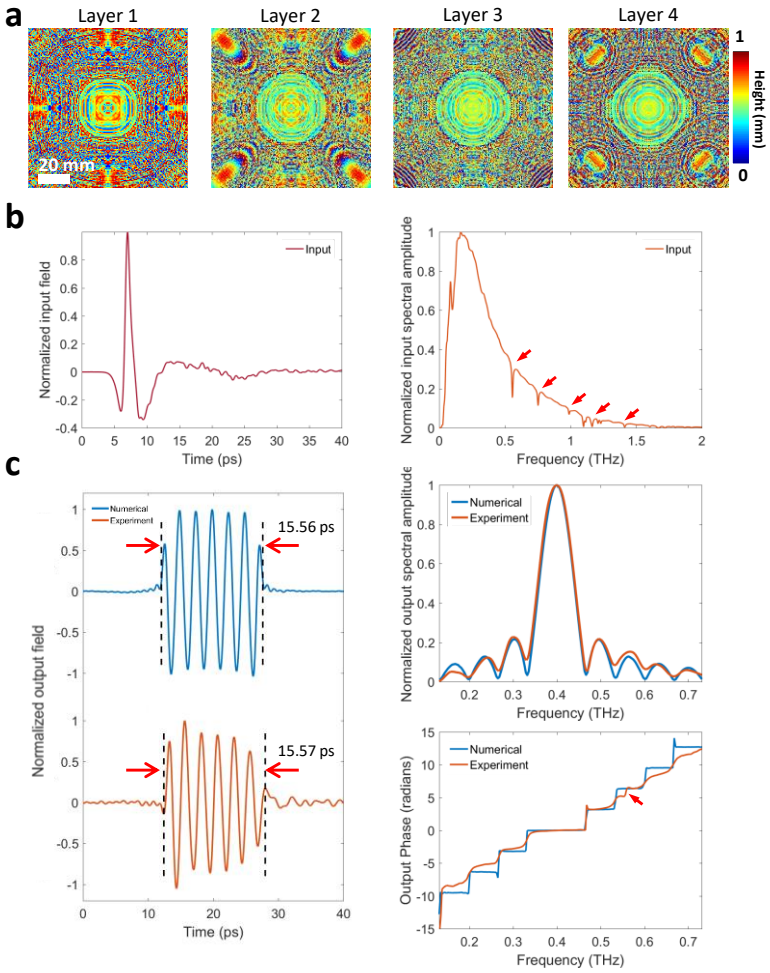

Fig. 3.7 Generic pulse shaping diffractive network design and output results. a The thickness profiles of the resulting diffractive layers after deep learning-based training in a computer. These diffractive layers synthesize a square pulse with a width of 15.56 ps over the output aperture for an input pulse shown in **b**. **b** Normalized amplitude of the input terahertz pulse measured right after the input aperture (see Fig. 1 of main text); in time-domain (left) and spectral domain (right). The red arrows on the measured spectral amplitude profile represent the water absorption bands at terahertz frequencies. **c Left:** The numerically computed (blue) and the experimentally measured (orange) output pulses in time domain. **Top right:** The normalized spectral amplitudes corresponding to the numerically computed (blue) and the experimentally measured (orange) pulses. **Bottom right:** Unwrapped spectral phase distributions computed based on the numerical forward model (blue) and experimentally measured (orange) pulse. Red arrow on the phase profile exemplifies a small discrepancy between the numerical and the experimental results due to the water absorption bands at THz frequencies.

Figure 3.7c reports the time-domain amplitude of the output waveform numerically computed (blue) based on these trained diffractive layers and the experimentally measured temporal waveform (orange) along with the corresponding spectral amplitude and phase distributions. The synthesized pulse shape by the 3D-printed diffractive network closely matches the numerically computed waveform using our forward model, despite the water absorption bands that appear in our experimental results, illustrated by the red arrows in Fig. 3.7b. The power efficiency at the carrier frequency (f_0 = 400 GHz) of this diffractive network was experimentally measured as ~0.97%. Figure 3.9 further demonstrates three additional generic pulse shaping diffractive network models that were trained with a flat input spectrum and experimentally tested using our terahertz setup to achieve different square pulses, with pulse widths of 11.25 ps, 13.45 ps and 16.69 ps, respectively, demonstrating a very good match to their numerical counterparts. The numerically computed peak frequencies for these three different diffractive networks were 399.4 GHz, 396.1 GHz and 399.4 GHz, which were measured experimentally as 399.1 GHz, 402.2 GHz and 401.8 GHz, respectively. As we move towards higher optical frequencies beyond 0.6 THz, the experimental spectral amplitude distributions start to deviate from their numerically predicted counterparts. Considering that the maximum material thickness in our model is ~1mm, at higher optical frequencies corresponding to wavelengths below ~0.5 mm, the light may travel more than 2 wavelengths inside a diffractive feature (depending on the final trained model) which will then violate the thin modulation layer assumption in our forward model contributing to some of the experimental errors observed in Fig. 3.9. In addition, the size each diffractive feature corresponding to a unique complex-valued modulation per neuron (see Methods) was chosen to be 0.5 mm due to the limited lateral resolution of our 3D printer. Therefore, for higher frequencies, the light fields are modulated at

each diffractive layer with 2D functions sampled at lower spatial rates, which, in return, partially limits the design capabilities of our diffractive networks at those smaller wavelengths of the pulse bandwidth. Furthermore, the uneven surface profile in 3D printing combined with thickness variations induced by fabrication imperfections contribute to some additional sources of experimental errors observed in our results.

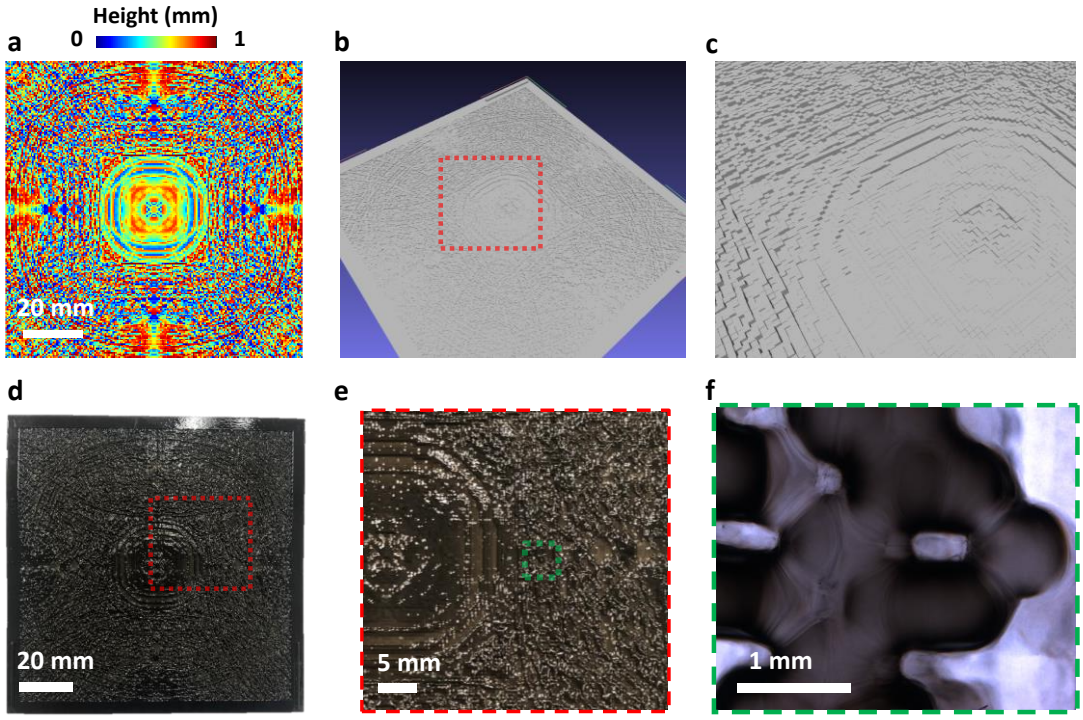

Fig. 3.8 Thickness profile of diffractive layer (a) The thickness profile of the first layer of the diffractive network reported in Fig. 3.7. **(b)** CAD drawing of the same layer and **(c)** zoomed-in version of the red square shown in (b). **(d)** The photo of the layer presented in (a), **(e)** zoomed-in version of the red square shown in (d) and **(f)** a microscopic image showing the fabricated structural details of the green square in (e).

To further demonstrate the design capabilities of our diffractive pulse shaping framework, in addition to the square pulses with various temporal widths reported earlier, we also trained three new diffractive network models that were designed to output (1) a chirped-Gaussian pulse (Fig.

3.10), (2) a sequence of positive and negative chirped Gaussian pulses, one following another (Fig. 3.11) and (3) a sequence of two chirp-free Gaussian pulses (Fig. 3.12). These figures report a very good match between the target, ground-truth pulse profiles and the output pulses synthesized by the corresponding trained diffractive network, both in time and spectral domains, clearly demonstrating the versatile nature of the presented framework to synthesize arbitrary pulses, shaped through the deep learning-based design of diffractive surfaces.

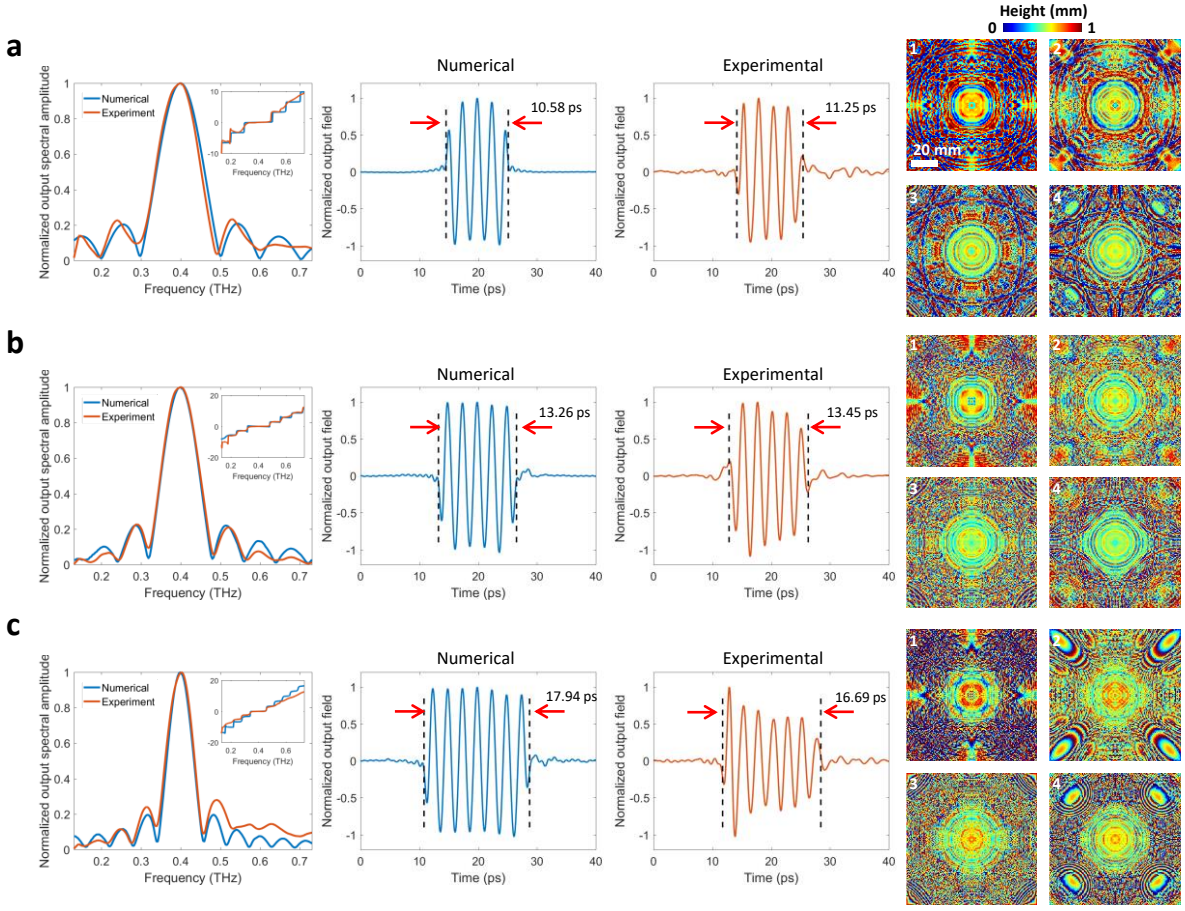

Fig. 3.9 Experimental validation of different generic pulse shaping diffractive networks. From left to right, the numerically computed (blue) and the experimentally measured (orange) normalized spectral amplitudes are illustrated with the inset plots showing the corresponding unwrapped spectral phase profiles; the numerically calculated (blue) normalized output pulse and the experimentally measured (orange) normalized output pulse are also shown along with the thickness profiles of the diffractive layers resulting from deep learning-based training for synthesizing the desired (ground truth) square pulses with pulse-widths of **(a)** 10.52 ps, **(b)** 13.02 ps and **(c)** 17.98 ps.

Next, we demonstrated the temporal width tunability of pulse shaping diffractive networks despite the passive nature of their layers. By changing the axial distance between successive diffractive layers by ΔZ, the temporal width and the peak frequency of the output waveform can

be tuned without any further training or a change to the 3D printed diffractive layers. We demonstrated this pulse-width tunability using the 3D printed diffractive network depicted in Fig. 3.7, but a similar tunability also applies to the network models shown in Fig. 3.9. Since our diffractive networks used 30 mm layer-to-layer distance in their design, we considered the ΔZ range to be between -10 mm to 20 mm; for instance, when ΔZ is taken as -10 mm, the axial distance between all the successive layers of the diffractive network is set to be 20 mm. Within this axial tuning range, Figs. 3.13a-h demonstrate the effect of changing this layer-to-layer distance of an already designed/trained diffractive network on the output waveform and its complex-valued spectrum. The results reveal that as the diffractive layers get closer to each other axially, i.e., a negative ΔZ, the pulse-width of the output waveform increases and the peak frequency decreases. For instance, when the axial distance between each diffractive layer of the design shown in Fig. 3.7 is decreased by 5 mm (ΔZ = -5 mm) as shown in Fig. 3.13d, the peak of the spectral amplitude distribution shifts from 399.4 GHz to 349.1 GHz according to our numerical forward model. The pulse-width of the resulting square pulse at the output aperture was numerically found to be 17.59 ps suggesting a longer pulse compared to 15.56 ps synthesized by the original design, ΔZ = 0 mm (Fig. 3.13d). The experimentally measured pulse width with the same amount of axial change in the layer-to-layer distance of the diffractive network revealed a 17.56 ps pulse after the spectral normalization step, confirming the tunability of our pulse shaping diffractive network and also providing a very good match to our numerical results (Fig. 3.13).

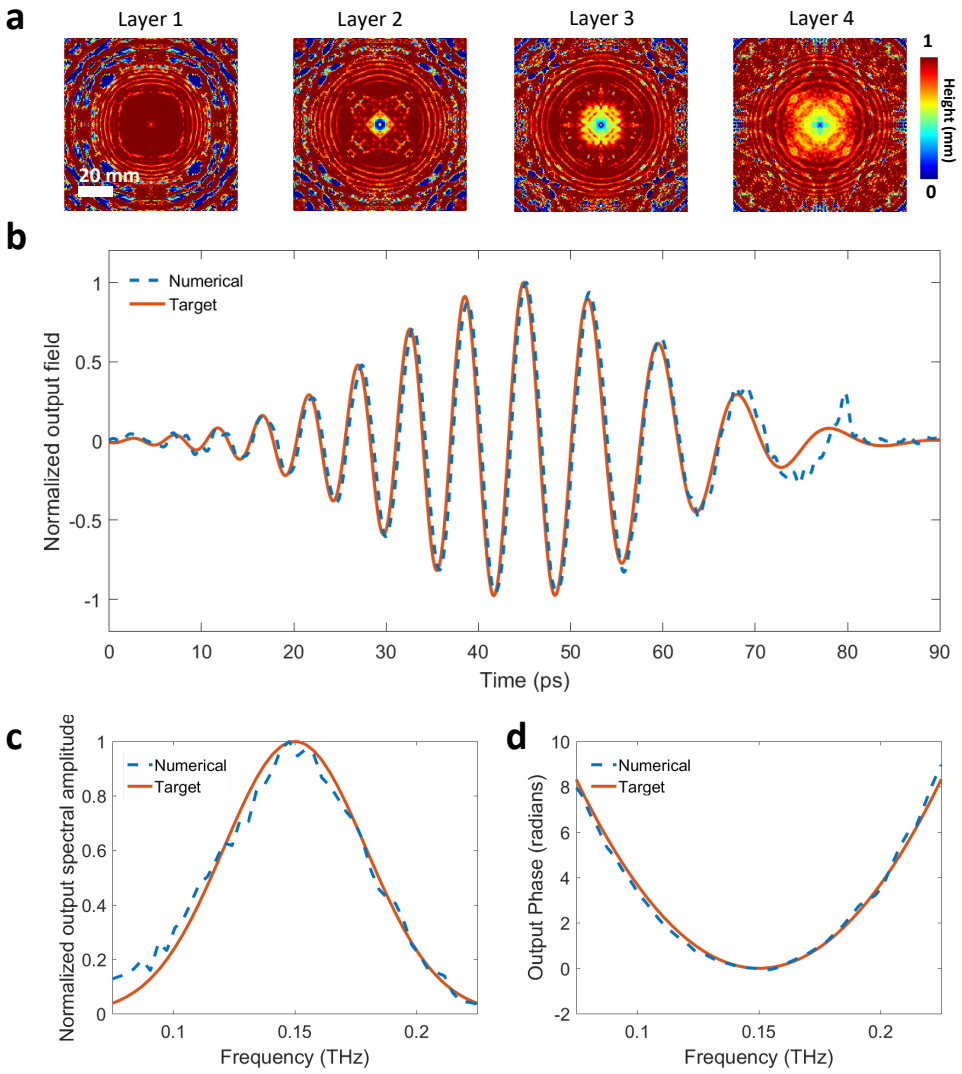

Fig. 3.10 Generic pulse shaping diffractive network design trained for synthesizing a chirped Gaussian pulse. **a** The thickness profiles of the resulting diffractive layers after deep learning-based training in a computer. These diffractive layers synthesize a chirped Gaussian pulse over the output aperture of the diffractive network. **b** The numerically computed (dashed blue) and the targeted ground-truth (orange) output pulses in time domain. **c** The normalized spectral amplitudes corresponding to the numerically computed (dashed blue) and the target (orange) pulses. **d** Unwrapped spectral phase distributions computed based on the numerical forward model (dashed blue) and the target (orange) pulse.

When the layer-to-layer distance is increased, i.e., a positive ΔZ, the output square pulse gets narrower in time domain with an accompanying shift in the peak frequency toward higher values. Figure 3.13e demonstrates an example of this case with ΔZ = 5 mm, i.e. the distance between each diffractive layer is increased to 35 mm. In this case, the experimentally measured and numerically computed square pulses at the output plane have peak frequencies of 451.4 GHz and 453.1 GHz, with the corresponding pulse-widths of 14.3 ps and 13.97 ps, respectively, once again confirming the tunability of our pulse shaping diffractive networks and demonstrating a very good agreement between the numerical forward model and our experiments. As we further increase ΔZ beyond 10 mm (depicted in Fig. 3.13f), the time domain pulse continues to get narrower.

To further explore methods to alter a given fabricated diffractive network and its output function, next we employed a Lego-like physical transfer learning approach to demonstrate pulse-width tunability by updating only part of a pre-trained network with newly trained and fabricated diffractive layers, showing the *modularity* of a diffractive pulse shaping network. For this aim, we took the pre-trained network that experimentally synthesized a 15.57 ps square waveform, noted as the original design in Fig. 3.14a, and further trained *only* the last diffractive layer to synthesize a new desired output waveform, i.e., a 12.03 ps square pulse, by keeping the first three layers as they are (already fabricated). We experimentally validated this transfer learning approach as shown in Fig. 3.14b by removing the existing last diffractive layer and inserting a newly trained layer, fabricated using the same 3D printer. Numerical and experimental results revealed very good match to each other for the normalized output spectral amplitude over a wide frequency range as well as for the normalized output field waveform, generating pulse-widths of 12.21 ps and 13.25 ps, respectively. Next we took an alternative

approach: this time, the last two diffractive layers were replaced with new diffractive layers trained to generate 12.03 ps square pulses. As illustrated in Fig. 3.14c, with the addition of these

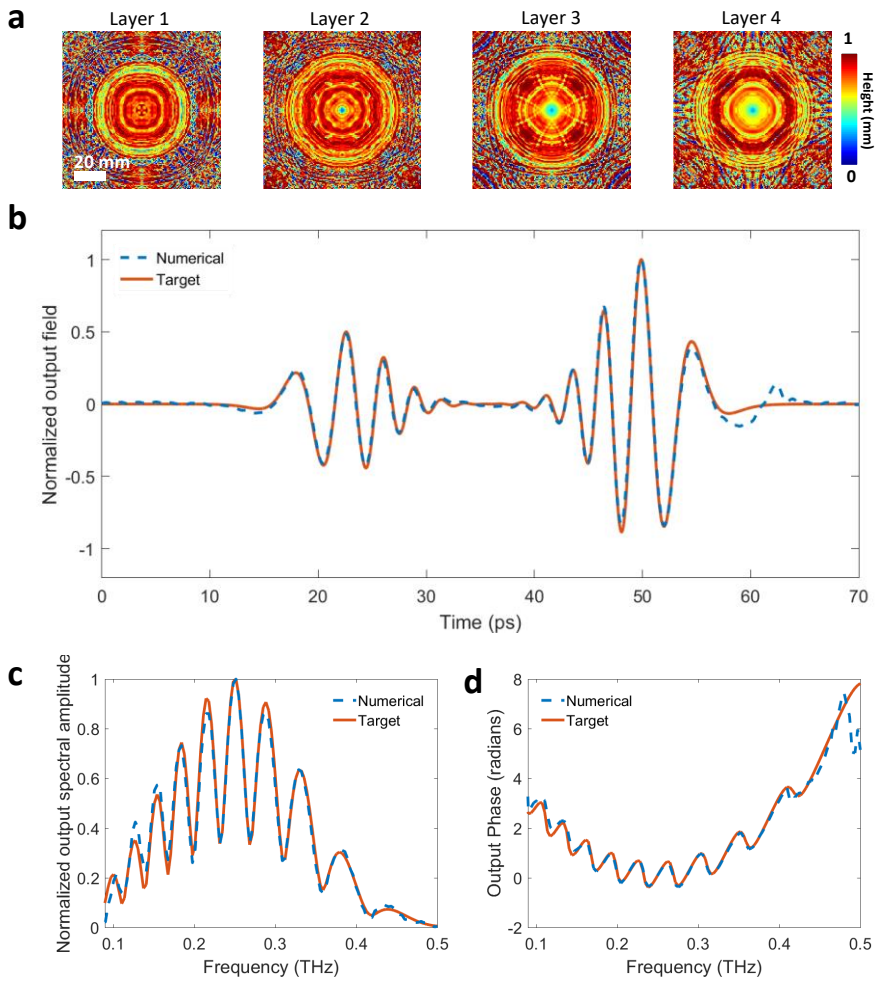

Fig. 3.11 Generic pulse shaping diffractive network design trained for synthesizing a sequence of positive and negative chirped Gaussian pulses. a The thickness profiles of the resulting diffractive layers after deep learning-based training in a computer. These diffractive layers synthesize a sequence of positive and negative chirped Gaussian pulses over the output aperture of the diffractive network. **b** The numerically computed (dashed blue) and the targeted ground-truth (orange) output pulses in time domain. **c** The normalized spectral amplitudes corresponding to the numerically computed (dashed blue) and the target (orange) pulses. **d** Unwrapped spectral phase distributions computed based on the numerical forward model (dashed blue) and target (orange) pulse.

two new diffractive layers to the already existing first two layers, the resulting new diffractive network successfully demonstrated the synthesis of 12.14 ps and 12.39 ps waveforms at the output aperture for the numerical and experimental waveforms, respectively. The peak frequency of the new network model was calculated to be at 399.4 GHz and it was experimentally measured to be at 399.8 GHz, showing once again a very good match between our numerical forward model and experimental results. Overall, the insertion of two newly trained layers, when compared to a single newly trained layer added on top of the existing layers of a 3D-fabricated network, provided us improved performance for achieving the new pulse form that is desired.

3.3 Discussion

The intrinsic pulse-width tunability of a given diffractive network that is achieved by changing the axial layer-to-layer distance is an interesting feature that we demonstrated numerically and experimentally: Figure 3.13a shows various pulse-widths obtained at seven different layer-to-layer distances using an existing network design. As the layer-to-layer distance of a diffractive network design increases, the temporal pulse-width at the output aperture gets smaller, without any further training or fabrication of new diffractive layers. This opens up the opportunity to synthesize new waveforms within a certain time window around the originally designed output pulse. In addition to that, an axial distance change between the existing layers of a diffractive network also shifts the center frequency of the output pulse as shown Figure 3.13b. As the diffractive layers get closer to each other, we observed a red-shift in the center frequency. Another related aspect of this pulse shaping diffractive framework is its modularity to tune the output pulses using a Lego-like physical transfer learning approach. By training a new layer (or layers) to replace part of an existing, pre-trained diffractive network model, on demand synthesis of new pulses can be achieved, as demonstrated in Figures 3.14b-c. These results highlight some

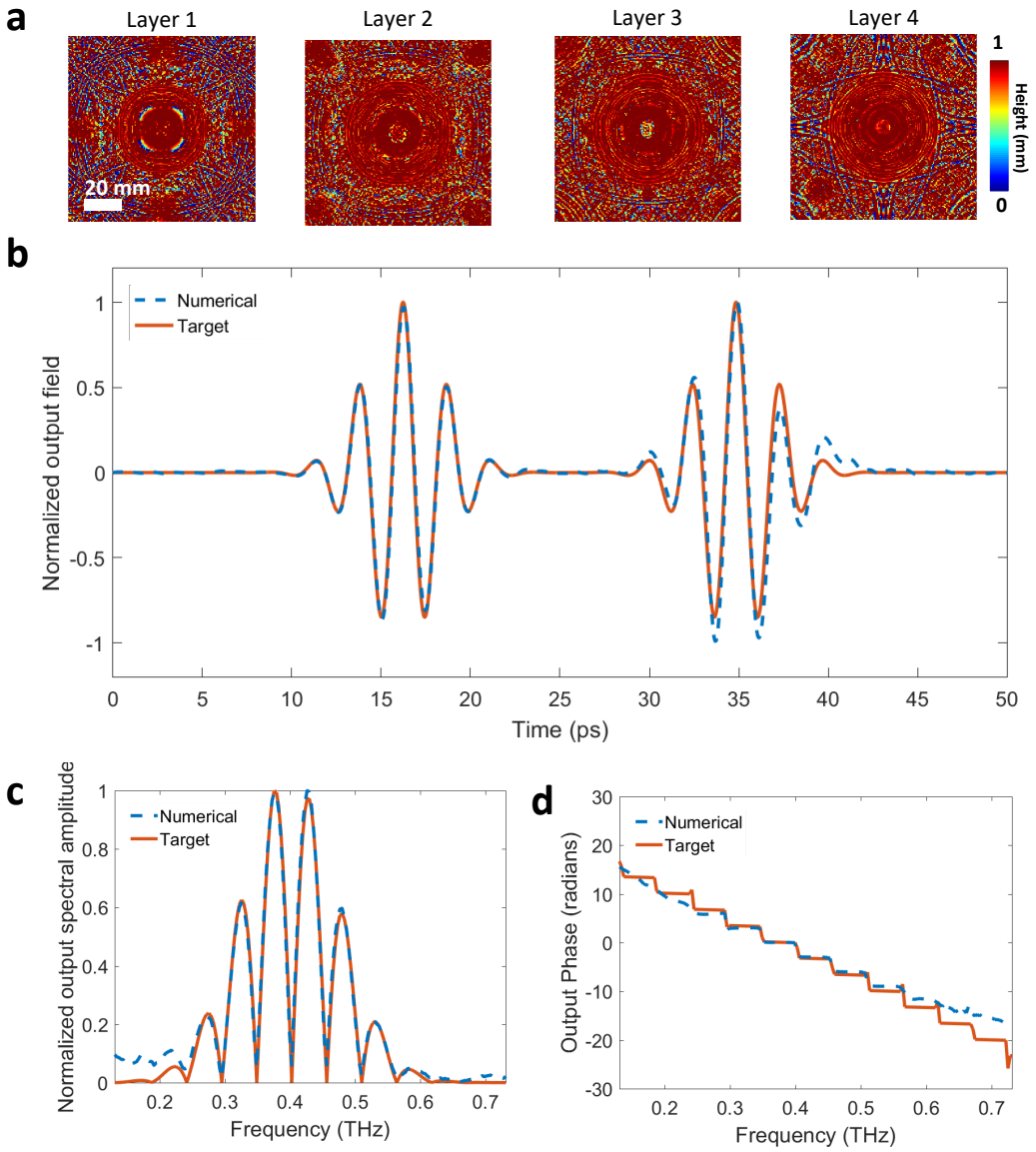

Fig. 3.12 Generic pulse shaping diffractive network design trained for synthesizing a sequence of chirp-free Gaussian pulses. a The thickness profiles of the resulting diffractive layers after deep learning-based training in a computer. These diffractive layers synthesize a sequence of chirp-free Gaussian pulses over the output aperture of the diffractive network**. b** The numerically computed (dashed blue) and targeted ground-truth (orange) output pulses in time domain. **c** The normalized spectral amplitudes corresponding to the numerically computed (dashed blue) and the target (orange) pulses. **d** Unwrapped spectral phase distributions computed based on the numerical forward model (dashed blue) and the target (orange) pulse.

of the unique features of diffractive pulse shaping networks and how they can adapt to potential changes in the desired output pulse patterns.

The presented pulse shaping framework has a compact design, with an axial length of approximately $250 \times \lambda_0$, where λ_0 denotes the peak wavelength. Moreover, it does not utilize any conventional optical components such as spatial light modulators, which makes it ideal for pulse shaping in terahertz part of the spectrum, where high-resolution spatio-temporal modulation and control of complex wavefronts over a broad bandwidth represent a significant challenge. In addition to being compact and much simpler compared to previous demonstrations of pulse shaping in terahertz spectrum, our results present the first implementation of direct pulse shaping in terahertz band, where the learned complex-valued spectral modulation function of the diffractive network directly acts on terahertz frequencies for pulse engineering. This capability enables new opportunities: when merged with appropriate fabrication methods and materials, the presented pulse shaping approach can be used to directly engineer terahertz pulses generated through quantum cascade lasers, solid-state circuits and particle accelerators. Another major advantage of this deep learning-based approach is that it is versatile and can be easily adapted to engineer terahertz pulses irrespective of their polarization state, beam quality as well as spectral/spatial aberrations.

The experimentally measured power efficiency values reported in our manuscript are ~1%. However, there exist various design strategies that can enable increased power-efficiency in diffractive pulse shaping networks as detailed in Table 3.1 (also see the Methods section). The diffractive networks reported in Table 3.1 were trained to synthesize 15.5 ps square pulses at their output plane. As one can observe in Table 3.1, the power efficiency values of the resulting

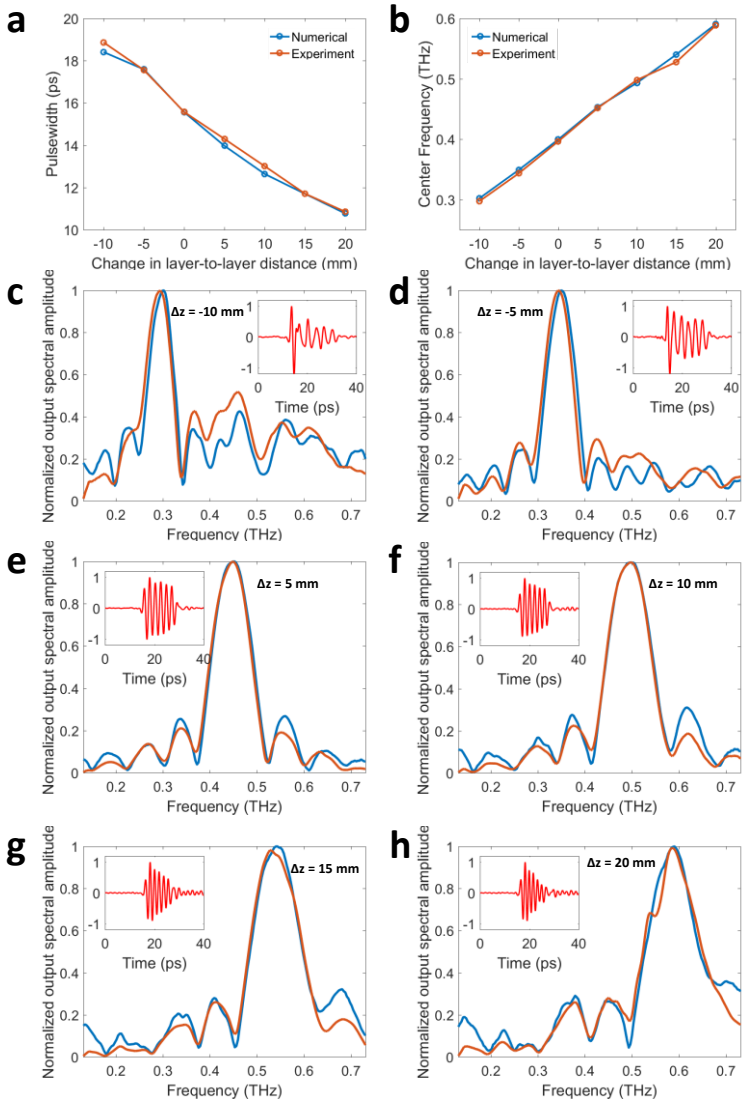

Fig. 3.13 Pulse width tunability of diffractive networks. (a) Numerically calculated and experimentally measured temporal pulse widths and **(b)** the corresponding shifts in the center frequency are depicted as a function of the inter-layer distances of a pulse shaping diffractive network that was originally trained for synthesizing a square pulse width of 15.50 ps (Δz = 0 mm, see Fig. 2). **(c-h)** The numerically computed (blue) and the experimentally measured (orange) normalized spectral amplitudes, with the inset plots showing the experimentally measured temporal waveform (red) when the layer-to-layer distances are changed by **(c)** Δz = -10 mm, **(d)** Δz = -5 mm, **(e)** Δz = 5 mm, **(f)** Δz = 10 mm, **(g)** Δz = 15 mm and **(h)** Δz = 20 mm. The negative (positive) sign indicates that the inter-layer axial distances decrease (increase).

diffractive models can be increased by more than an order of magnitude by adjusting the training loss function, increasing the output aperture size and using low absorption materials. For example, as reported in the second column of Table 3.1, when the material absorption is ignored during the testing of a diffractive network model, a 2-fold wider output aperture (i.e., 4 mm) provides a significant improvement in the power efficiency of the pulse shaping networks, reaching 60.37% and 61% for two different network models. On the other hand, if the absorption of our 3D-printing material is taken into account as part of the optical forward model, one can reach an efficiency value of 17.84% by accordingly optimizing the training loss function and using a 4 mm output aperture (see Table 3.1).

By comparing the top and bottom efficiency values for a given training loss function and design strategy reported in Table 3.1, we clearly see that the 3D-printing material used in this work decreases the pulse shaping network efficiency 2-5 times, in different designs, compared to an ideal, non-absorbing optical material. As an alternative fabrication material for diffractive pulse shaping networks, one can consider low-absorption polymers[134–136] used in commercially available components designed for THz wavelengths, e.g., TPX, which exhibits a two orders-of-magnitude smaller absorption coefficient compared to the 3D printing material used in our work. There have been various fabrication processes developed for such low absorption polymers[137,138], which can be used to precisely control the thickness of these low-loss polymers with a relatively high-resolution (~$\lambda/2$) to manufacture pulse shaping diffractive networks with much lower material absorption. To even further improve the output efficiency of pulse shaping diffractive networks, anti-reflective (AR) coatings over diffractive surfaces can also be utilized to reduce back-reflections, similar to the AR-coated commercial lenses and other optical components.

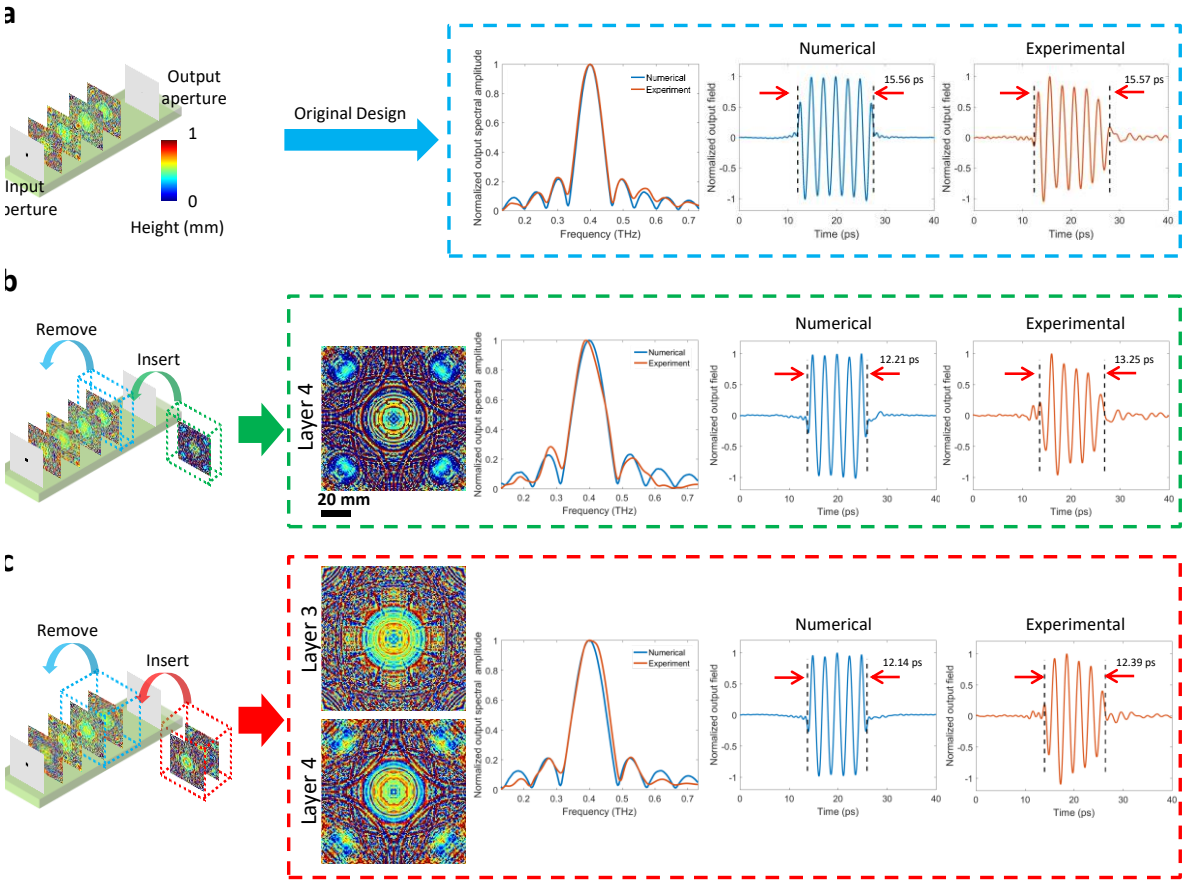

Fig. 3.14 Changing the output temporal waveform of a diffractive network by a Lego-like transfer learning approach. **a** The temporal and spectral output distributions (blue dashed box), synthesized by the original diffractive design that was trained to generate a 15.50 ps square pulse. **b** Replacing the last diffractive layer with another, newly trained diffractive layer to synthesize a 12.03 ps square pulse at the output. The thickness profile of the newly trained diffractive layer is shown, together with the normalized spectral and temporal profiles synthesized by this new diffractive network in green dashed box. **c** Replacing the last two diffractive layers with newly trained, two diffractive layers to synthesize a 12.03 ps square pulse at the output. The thickness profiles of the newly trained diffractive layers are shown, together with the normalized spectral and temporal profiles synthesized by this new diffractive network in red dashed box.

In conclusion, we presented a modular pulse shaping network that synthesizes various pulse waveforms using deep learning. Precise shaping of the spectral amplitude and phase profile of an arbitrary input pulse over a wide frequency range can be achieved using this platform, which will be transformative for various applications including e.g., communications, pulse compression, ultra-fast imaging and spectroscopy. In addition to direct engineering of terahertz pulses, the presented diffractive pulse shaping network can be utilized in different parts of the electromagnetic spectrum by using appropriate fabrication technologies and materials.

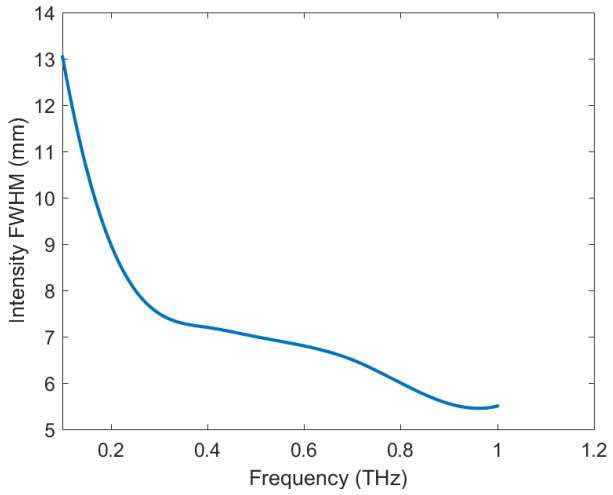

Fig. 3.15 Experimentally measured full width at half maximum (FWHM) values of the spatial intensity profiles of different spectral components in the THz beam at the input aperture plane of the diffractive network.

3.4 Methods

Terahertz setup

Figure 3.1 shows the schematic diagram of the terahertz time-domain spectroscopy (THz-TDS) setup that was used to measure the input and output pulse profiles reported in this work. A Ti:sapphire laser (Coherent Mira HP) is used to generate femtosecond optical pulses. The optical

beam generated by the laser is split into two parts. One part of the beam is used to pump a high-power plasmonic photoconductive terahertz source to generate terahertz pulses[139], which are collimated with off-axis parabolic mirrors and guided to a high-sensitivity plasmonic photoconductive terahertz detector[133]. The other part of the beam passes through an optical delay line (Newport IMS300LM) and is focused onto the terahertz detector. As a result, an ultrafast signal which is directly proportional to the incident terahertz field is generated within the terahertz detector. The signal is sampled with a 12.5 fs time-resolution over a 400 ps time-window by changing the time delay between the terahertz and optical probe pulses incident on the detector, amplified with a transimpedance pre-amplifier (Femto DHPCA-100), and acquired with a lock-in amplifier (Zurich Instruments MFLI). For each measurement, 10 time-domain traces are collected and averaged. The described THz-TDS setup provides a 90 dB signal-to-noise ratio over a 5 THz noise-equivalent-power bandwidth.

Each one of the pulse shaping diffractive networks consists of 4 trained layers that are separated by 3 cm as illustrated in Fig. 3.1. The diffractive layers, input and output apertures, were fabricated using a 3D Printer (Objet30 Pro, Stratasys Ltd.). The fabrication/preparation of each diffractive layer takes approximately 1.5-2 hours. A square input aperture (0.8 cm) and an output aperture (0.2 cm) are placed 3 cm from the first diffractive layer and 10 cm from the last diffractive layer, respectively (Fig. 3.1c). The printed apertures were aluminum coated to prevent any light wave passing through the regions outside of the aperture. After the design and printing of the diffractive layers, they were placed at their corresponding locations inside a 3D printed holder that ensures robust alignment between the layers. During the pulse shaping experiments, the diffractive network was directly placed between the terahertz source and detector, coaxial

		2 mm width output aperture	4 mm width output aperture
Trained without absorption $L_p = \frac{\sum_\omega (I_{target} - I_{sensor})^2}{\sum_\omega I_{target}^2}$		**Efficiency**	**Efficiency**
	Test without absorption	**14.91%**	**60.37%**
	Test with absorption	**3.9%**	**12%**
Trained with absorption $L_p = \frac{\sum_\omega (I_{target} - I_{sensor})^2}{\sum_\omega I_{target}^2}$	Test without absorption	**46.13%**	**61%**
	Test with absorption	**13.13%**	**15.78%**
Trained with absorption $L_p = \begin{cases} -\log\left(\frac{\eta}{\eta_{th}}\right), & \text{if } \eta < \eta_{th} \\ 0, & \text{if } \eta \geq \eta_{th} \end{cases}$	Test without absorption	**31.14%**	**31.93%**
	Test with absorption	**14.11%**	**17.84%**

Table 3.1 Power efficiency values of pulse shaping diffractive networks trained with different loss functions.

with the terahertz input pulse emanating from the source (Figs. 3.1b,d). After the alignment of the diffractive network, the output pulse was measured and it was followed by the measurement of the reference input pulse which was acquired by placing the same terahertz detector at the input aperture, without any diffractive layers between the source and detector. For generic diffractive networks that were trained with flat input spectra, the measured output pulse spectrum is normalized with respect to the measured reference input pulse and its spectral amplitude is smoothened around water absorption lines shown in Figs. 3.7, 3.9, 3.13 and 3.14. The measured pulse width at the network output is defined as the width of the time interval that the envelope of the pulse amplitude is at least 20% of its maximum.

Forward Model

Our forward model considers the layers of a diffractive network as thin modulation elements that are connected to the next layer through free space propagation. The modulation of neurons at each layer can be modeled as:

$$M^n(x_i, y_i, z_i, \lambda) = A^n(x_i, y_i, z_i, \lambda) \exp\left(j\varphi^n(x_i, y_i, z_i, \lambda)\right) \quad (3.1),$$

where M represents the complex transmission/reflection coefficient. The field amplitude, phase, wavelength, and diffractive layer number are denoted by A, ϕ, λ and n respectively. Free space propagation between each layer is calculated based on the Rayleigh-Sommerfeld formulation of diffraction that models a diffractive feature as source of a secondary wave:

$$W_i^n(x, y, z, \lambda) = \frac{z - z_i}{r^2}\left(\frac{1}{2\pi r} + \frac{1}{j\lambda}\right) \exp\left(\frac{j2\pi r}{\lambda}\right) \quad (3.2),$$

where $r = \sqrt{(x-x_i)^2 + (y-y_i)^2 + (z-z_i)^2}$, $j = \sqrt{-1}$ and $W_i^n(x, y, z, \lambda)$ is the secondary wave generated by the i[th] neuron on n[th] layer at location (x_i, y_i, z_i), respectively. Then, we can write the optical field at layer n, at point (x_i, y_i, z_i) as:

$$U^n(x_i, y_i, z_i, \lambda) = M^n(x_i, y_i, z_i, \lambda) \sum_k U^{n-1}(x_k, y_k, z_k, \lambda) W_k^{n-1}(x_i, y_i, z_i, \lambda), n \geq 1 \quad (3.3).$$

Network Training

During the training of a pulse shaping diffractive network, one of the 5 pulses measured at the input plane (Figs. 3.4a-b) were randomly selected as the input pulse at each iteration of the training model for the diffractive networks reported in Figs. 3.2 and 3.3; for the generic diffractive network models reported in Fig. 3.7 and Figs. 3.9-3.14, however, the input is modeled as a spectrally flat Gaussian beam with varying FWHM values over a wide frequency range (Fig. 3.15) and with a uniform phase profile. The wave propagation is performed for N = 300 discrete frequencies that were uniformly sampled between 3 GHz and 1 THz.

In our wave propagation through the diffractive layers, a 0.5 mm pixel (i.e., diffractive feature) size is assumed based on the lateral resolution of our 3D printer. While a pixel size of 0.5 mm can create all the propagating modes of free-space for frequencies below ~300 GHz, they can only excite plane waves over a subset of the k-vectors supported by the free-space for the spectral components between 300 GHz and 1 THz.[140] Therefore, our diffractive pulse shaping networks would benefit from higher resolution fabrication techniques with better lateral resolution to more accurately control and engineer the complex-valued spectral weights of a given desired pulse.

To calculate the Rayleigh-Sommerfeld integral more accurately, each pixel is oversampled twice so that all 4 elements have the same thickness values in that 2×2 grid. The thickness of each pixel, h, is composed of a base height (h_{base}) of 0.1 mm, which provides adequate mechanical stiffness to the fabricated diffractive layer and a trainable modulation height (h_{tr}) that is between 0 and 1 mm, i.e.,

$$h = h_{base} + h_{tr} \qquad (3.4)$$

To confine the modulation height between 0 and 1 mm, we defined h_{tr} over an auxiliary training-related variable, h_a, using:

$$h_{tr} = 0.5mm \times \{1 + \sin(h_a)\} \qquad (3.5),$$

In its general form, the amplitude and phase modulation of each neuron of a given diffractive layer is a function of the layer thickness, incident wavelength, material extinction coefficient $\kappa(\lambda)$ and refractive index $n(\lambda)$, i.e.,

$$A^n(x, y, z, \lambda) = \exp\left(-\frac{2\pi\kappa(\lambda)h}{\lambda}\right) \qquad (3.6)$$

$$\varphi^n(x, y, z, \lambda) = \frac{2\pi h(n(\lambda) - n_{air})}{\lambda} \qquad (3.7)$$

The material refractive index $n(\lambda)$ and the extinction coefficient $\kappa(\lambda)$ are defined as the real and imaginary parts of the complex refractive index, $\tilde{n}(\lambda) = n(\lambda) + j\kappa(\lambda)$, determined by the dispersion of our 3D fabrication material[92]. Since we have relatively small variations in the extinction coefficient over the frequency band that we utilized in this work, we ignored the material absorption during the training and numerical simulations of diffractive layers.

After the wave propagation through diffractive layers, light goes through the output square aperture of 2mm width, which is placed right in front of the hemisphere silicon lens which is 1.2 cm in diameter. Since the effective aperture of this Si lens was significantly restricted by the output aperture, it was modeled as a uniform slab with a refractive index of 3.4 and 0.5 cm thickness. After the propagation through the Si slab, the coherent integration of the optical waves incident on the active area of the detector was computed to obtain the spectral field amplitude and phase for each frequency. The power efficiency was defined as $\eta_{f_0} = \frac{I_{sensor,f_0}}{I_{input,f_0}}$ for the peak/center frequency (f_0) of given diffractive network design, where I_{input,f_0} and I_{sensor,f_0} denote the power within the input and output apertures, respectively.

Our loss function (L) used during the training phase has three components: temporal loss term (L_t) which penalizes the mismatch between the target and the output time waveforms, the power loss term (L_p), and the power surrounding the detector region (L_s), i.e.,

$$L = \alpha L_t + \beta L_p + L_s \qquad (3.8)$$

To calculate the temporal loss, L_t, first the output temporal waveform is reconstructed from the spectral field amplitude and phase on the detector area, and it is normalized. Then, the difference between the target temporal waveform and the reconstructed output waveform is integrated over time:

$$L_t = \sum_t \left(f_{target} - f_{output}\right)^2 \qquad (3.9),$$

where f_{target} and f_{output} denote the ground-truth, time-domain waveform and the synthesized waveform by the diffractive network model at a training iteration. For a given

diffractive network model, f_{output} is computed by propagating the input waves of all the spectral components from the input aperture to the output aperture. Next, the complex-valued wave fields of these different wavelength components are integrated over the sensitive area of the detector to obtain each complex-valued spectral coefficient at the output, which is followed by an inverse Fourier transform operation over the resulting vector. Alternatively, the error term between a target, time-domain pulse, f_{target}, and the synthesized waveform by the diffractive network, f_{output}, can directly be computed based on the complex-valued spectral coefficients without any inverse Fourier transform operation. However, in this case, since the error is defined based on the complex-valued target and output functions, two separate error functions must be computed for the real and imaginary parts of the spectral coefficients and these two losses must be combined to compute the final loss term.

The analytical form of the square pulses used in this work can be written as: $f_{target}(t) = rect(bt)\cos(2\pi f_0 t)$, where f_0 and b represent the carrier frequency and the rectangular pulse-width, respectively. For the Gaussian pulses, however, the analytical form of the target waveform can be written as: $f_{target}(t) = \sum_{i=1}^{n} \mathrm{C_i}\cos\left(2\pi f_0 (t - t_{0,i})\right) \exp\left(-\left(t - t_{0,i}\right)^2 / (2p_i)\right) \exp\left(jq_i (t - t_{0,i})^2\right)$, where $t_{0,i}$, C_i, p_i and q_i denote the time instant of the peak, magnitude, variance of the low-pass envelope and the instantaneous angular chirpiness, respectively. The number of desired pulses inside a targeted time-window is determined by n. For the three examples shown in Figs. 3.10-3.12, the target time domain waveforms were created by setting these parameters to [n = 1, t_0 = 0, C_1 = 1, p_1 = 2.2×10^{-22}, q_1 = 5.76×10^{21}]; [n = 2, $t_{0,1}$ = 0, $t_{0,2}$ = 27 ps, C_1 = 1, C_2 = 0.5, p_1 = p_2 =1.38×10^{-23}, q_1 = 6.25×10^{22} q_2 = -6.25×10^{22}]; and [n = 2, $t_{0,1}$ = 0, $t_{0,2}$ = 19 ps, C_1 = 1, C_2 = 1, p_1 = p_2 =4.58×10^{-24}, q_1 = q_2 = 0], respectively.

For the diffractive network designs shown in last row of Table 3.1, we used a power loss term, L_p, defined as:

$$L_p = \begin{cases} -\log\left(\frac{\eta}{\eta_{\text{th}}}\right), & \text{if } \eta < \eta_{\text{th}} \\ 0, & \text{if } \eta \geq \eta_{\text{th}} \end{cases} \qquad (3.10),$$

where $= \frac{\sum_\omega I_{sensor}}{\sum_\omega I_{input}}$. I_{input} and I_{sensor} denote the power within the input and output apertures for a given wavelength, respectively. For the diffractive network designs shown in Table 3.1 last row, corresponding to 2mm and 4mm output apertures, η_{th} was selected as 0.07 and 0.08, respectively. For the all remaining designs reported in the manuscript, the power loss term is defined as:

$$L_p = \frac{\sum_\omega \left(I_{target} - I_{sensor}\right)^2}{\sum_\omega I_{target}^2} \qquad (3.11),$$

where I_{target} is the total power of the target waveform at a given wavelength within the input aperture, normalized with respect to the power of the input at the center frequency, f_0.

The last component of our loss function which represents the power surrounding the detector aperture is defined as:

$$L_s = \frac{\sum_\omega I_{surround}}{\sum_\omega I_{output\ plane}} \qquad (3.12),$$

where $I_{surround}$ is the total power at a given wavelength within the 5 mm × 5 mm square region that is centered around the output aperture (excluding the output aperture, i.e., it only measures the signal surrounding the output aperture) and $I_{output\ plane}$ is the total power at a given wavelength within the output plane.

The diffractive networks that synthesized 10.58 ps, 10.96 ps, 13.26 ps, 15.56 ps, 15.69 ps and 17.94 ps square terahertz pulses were trained with $\frac{\alpha}{\beta}$ ratios of 6500, 500, 4500, 1500, 750000 and 2500, respectively. For the Lego-like transfer learning approach, an $\frac{\alpha}{\beta}$ ratio of 8500 was used. For Figs. 3.10 and 3.12, the forward optical model was trained using an $\frac{\alpha}{\beta}$ ratio of 1500 and for Fig. 3.11, we used $\frac{\alpha}{\beta} = 15000$.

Table 3.1 reports a series of diffractive optical network designs that are trained to create a square pulse of 15.5 ps at their output apertures, achieving different levels of power efficiencies. Among these pulse shaping diffractive network models, the $\frac{\alpha}{\beta}$ ratio was adjusted depending on the size of the output aperture. Specifically, the diffractive networks targeting a 2mm aperture at the output plane were trained with $\frac{\alpha}{\beta} = 1500$, and this ratio was reduced to 136 for the diffractive pulse shaping systems with 4mm wide output apertures. Finally, an $\frac{\alpha}{\beta}$ ratio of 150 was used for the diffractive optical networks that were trained with the power efficiency loss term described in Eq. 3.10.

In our training, Adam optimizer is used as a standard error backpropagation method with a learning rate of 0.8×10^{-3} for chirped-Gaussian pulses depicted in Figs. 3.10-3.11 and the chirp-free Gaussian pulse train reported in Fig. 3.12. For the diffractive networks synthesizing 10.96 ps and 15.69 ps square pulses, on the other hand, the learning rate was set to be 10^{-3}. For the rest of the diffractive network models 10^{-4} was used as the learning rate. All the trainable parameters were initialized as zero. Our designs used Python (v3.7.3) and Tensorflow (v1.15.0) on a computer that has Nvidia Titan RTX graphical processing unit, Intel Core i9 CPU and 128 GB of

RAM with Windows 10 operating system. MATLAB 2016b is used to convert designed diffractive layers to a 3D printable (.stl) file format.

Chapter 4 Conclusion

Machine learning has been revolutionizing many areas of optics, including optical sensors and devices. Recent advances in computational imaging and deep learning improved our capabilities by providing noise reduction and better sensitivity in optical sensors and also helped us to overcome non-intuitive design problems for optical devices.

In summary, my contributions to machine learning-enabled optical sensors are presented in Chapter 1. I introduced a platform to non-invasively analyze human ocular microbiome by computational imaging of contact lenses and machine learning based detection of bacteria, *Staphylococcus aureus*. This platform opens up an avenue for monitoring and detection of other bacteria in ocular microbiome with small changes in surface chemistry protocol, also enables further analysis and understanding between ocular microbiome and some diseases, by providing easy access to ocular environment.

In chapter 2, a new framework that is based on deep learning has been introduced for all-optical machine learning applications such as image classification and for solving non-inverse problems in optics. This diffractive optical network framework, which consists of passive layers, brings wave optics and deep learning together to tackle tasks with the speed of light processing, low power consumption while enabling parallel computing.

In chapter 3, I introduced optical pulse shaping platform which is based on diffractive optical surfaces to synthesize arbitrary waveform from input pulse. Deep learning based training enables us to solve such interesting non-intuitive optical problems without using any conventional optical components. First direct pulse shaping in terahertz spectrum was demonstrated by synthetizing

arbitrary temporal waveforms. This avenue shows potential opportunities for the design of new optical devices using deep learning based diffractive optical networks.

References

1 . Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett C, Knight R, Gordon JI. The human microbiome project: exploring the microbial part of ourselves in a changing world. *Nature* 2007; **449**: 804–810.

2 . Dong Q, Brulc JM, Iovieno A, Bates B, Garoutte A, Miller D *et al.* Diversity of Bacteria at Healthy Human Conjunctiva. *Invest Ophthalmol Vis Sci* 2011; **52**: 5408–5413.

3. Beck JM, Young VB, Huffnagle GB. The microbiome of the lung. *Translational Research* 2012; **160**: 258–266.

4. Consortium THMP. Structure, function and diversity of the healthy human microbiome. *Nature* 2012; **486**: 207–214.

5. Biteen JS, Blainey PC, Cardon ZG, Chun M, Church GM, Dorrestein PC *et al.* Tools for the Microbiome: Nano and Beyond. *ACS Nano* 2016; **10**: 6–37.

6. Dayan GH, Mohamed N, Scully IL, Cooper D, Begier E, Eiden J *et al.* Staphylococcus aureus: the current state of disease, pathophysiology and strategies for prevention. *Expert Review of Vaccines* 2016; **15**: 1373–1392.

7. Miller D, Iovieno A. The role of microbial flora on the ocular surface. *Curr Opin Allergy Clin Immunol* 2009; **9**: 466–470.

8. Larkin DFP, Leeming JP. Quantitative alterations of the commensal eye bacteria in contact lens wear. *Eye* 1991; **5**: 70–74.

9. Tong SYC, Davis JS, Eichenberger E, Holland TL, Fowler VG. Staphylococcus aureus Infections: Epidemiology, Pathophysiology, Clinical Manifestations, and Management. *Clin Microbiol Rev* 2015; **28**: 603–661.

10. Clooney AG, Fouhy F, Sleator RD, Driscoll AO, Stanton C, Cotter PD *et al.* Comparing Apples and Oranges?: Next Generation Sequencing and Its Impact on Microbiome Analysis. *PLOS ONE* 2016; **11**: e0148028.

11. Salter SJ, Cox MJ, Turek EM, Calus ST, Cookson WO, Moffatt MF *et al.* Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biology* 2014; **12**: 87.

12. Patel JB. 16S rRNA gene sequencing for bacterial pathogen identification in the clinical laboratory. *Molecular Diagnosis* 2001; **6**: 313–321.

13. Willcox MDP. Characterization of the normal microbiota of the ocular surface. *Experimental Eye Research* 2013; **117**: 99–105.

14. Mak WC, Cheung KY, Orban J, Lee C-J, Turner APF, Griffith M. Surface-Engineered Contact Lens as an Advanced Theranostic Platform for Modulation and Detection of Viral Infection. *ACS Appl Mater Interfaces* 2015; **7**: 25487–25494.

15. 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 Meth* 2012; **9**: 889–895.

16. Tseng D, Mudanyali O, Oztoprak C, Isikman SO, Sencan I, Yaglidere O *et al.* Lensfree microscopy on a cellphone. *Lab Chip* 2010; **10**: 1787–1792.

17. Ozcan A, McLeod E. Lensless Imaging and Sensing. *Annual Review of Biomedical Engineering* 2016; **18**: 77–102.

18. Göröcs Z, Ozcan A. On-Chip Biomedical Imaging. *IEEE Reviews in Biomedical Engineering* 2013; **6**: 29–46.

19. Decher G. Fuzzy Nanoassemblies: Toward Layered Polymeric Multicomposites. *Science* 1997; **277**: 1232–1237.

20. Hugo WB. *An Introduction to Microbiology: Pharmaceutical Monographs*. Butterworth-Heinemann, 2014.

21. Santos L, Rodrigues D, Lira M, Oliveira MECDR, Oliveira R, Vilar EY-P *et al.* The influence of surface treatment on hydrophobicity, protein adsorption and microbial colonisation of silicone hydrogel contact lenses. *Contact Lens and Anterior Eye* 2007; **30**: 183–188.

22. Feizi A, Zhang Y, Greenbaum A, Guziak A, Luong M, Lok Chan RY *et al.* Rapid, portable and cost-effective yeast cell viability and concentration analysis using lensfree on-chip microscopy and machine learning. *Lab on a Chip* 2016; **16**: 4350–4358.

23. Memmolo P, Paturzo M, Javidi B, Netti PA, Ferraro P. Refocusing criterion via sparsity measurements in digital holography. *Opt Lett, OL* 2014; **39**: 4719–4722.

24. Memmolo P, Miccio L, Paturzo M, Caprio GD, Coppola G, Netti PA *et al.* Recent advances in holographic 3D particle tracking. *Adv Opt Photon, AOP* 2015; **7**: 713–755.

25. Merola F, Miccio L, Memmolo P, Paturzo M, Grilli S, Ferraro P. Simultaneous Optical Manipulation, 3-D Tracking, and Imaging of Micro-Objects by Digital Holography in Microfluidics. *IEEE Photonics Journal* 2012; **4**: 451–454.

26. Wu Y-C, Shiledar A, Li Y-C, Wong J, Feng S, Chen X *et al.* Air quality monitoring using mobile microscopy and machine learning. *Light Sci Appl* 2017; **6**: e17046.

27. Zhang Y, Wang H, Wu Y, Tamamitsu M, Ozcan A. Edge sparsity criterion for robust holographic autofocusing. *Opt Lett, OL* 2017; **42**: 3824–3827.

28. Matsushima K. Formulation of the rotational transformation of wave fields and their application to digital holography. *Appl Opt, AO* 2008; **47**: D110–D116.

29. 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. *Science Translational Medicine* 2014; **6**: 267ra175-267ra175.

30. Ferraro P, Paturzo M, Memmolo P, Finizio A. Controlling depth of focus in 3D image reconstructions by flexible and adaptive deformation of digital holograms. *Opt Lett, OL* 2009; **34**: 2787–2789.

31. Paturzo M, Ferraro P. Creating an extended focus image of a tilted object in Fourier digital holography. *Opt Express, OE* 2009; **17**: 20546–20552.

32. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015; **521**: 436–444.

33. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M *et al.* A survey on deep learning in medical image analysis. *Medical Image Analysis* 2017; **42**: 60–88.

34. Graves A, Mohamed A, Hinton G. Speech recognition with deep recurrent neural networks. In: *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. 2013, pp 6645–6649.

35. Cho K, van Merrienboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H *et al.* Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. 2014.https://arxiv.org/abs/1406.1078v3 (accessed 2 Dec2020).

36. Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems* 2012; **25**: 1097–1105.

37. Silver D, Huang A, Maddison CJ, Guez A, Sifre L, van den Driessche G *et al.* Mastering the game of Go with deep neural networks and tree search. *Nature* 2016; **529**: 484–489.

38. Kamilov US, Papadopoulos IN, Shoreh MH, Goy A, Vonesch C, Unser M *et al.* Learning approach to optical tomography. *Optica, OPTICA* 2015; **2**: 517–522.

39. Rivenson Y, Göröcs Z, Günaydin H, Zhang Y, Wang H, Ozcan A. Deep learning microscopy. *Optica, OPTICA* 2017; **4**: 1437–1443.

40. Jin KH, McCann MT, Froustey E, Unser M. Deep Convolutional Neural Network for Inverse Problems in Imaging. *IEEE Transactions on Image Processing* 2017; **26**: 4509–4522.

41. 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.

42. Sinha A, Lee J, Li S, Barbastathis G. Lensless computational imaging through deep learning. *Optica, OPTICA* 2017; **4**: 1117–1125.

43. Hammernik K, Klatzer T, Kobler E, Recht MP, Sodickson DK, Pock T *et al.* Learning a variational network for reconstruction of accelerated MRI data. *Magnetic Resonance in Medicine* 2018; **79**: 3055–3071.

44. 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.

45. Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 1998; **86**: 2278–2324.

46. Ciregan D, Meier U, Schmidhuber J. Multi-column deep neural networks for image classification. In: *2012 IEEE Conference on Computer Vision and Pattern Recognition*. 2012, pp 3642–3649.

47. Lee C-Y, Gallagher PW, Tu Z. Generalizing Pooling Functions in Convolutional Neural Networks: Mixed, Gated, and Tree. ; : 9.

48. Ranzato MA, Poultney C, Chopra S, LeCun Y. Efficient learning of sparse representations with an energy-based model. In: *Advances in Neural Information Processing Systems 19 - Proceedings of the 2006 Conference*. 2007, pp 1137–1144.

49. 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 4 Dec2020).

50. Zeng S, Zhang B, Zhang Y, Gou J. Collaboratively Weighting Deep and Classic Representation via L2 Regularization for Image Classification. *arXiv:180207589 [cs]* 2018.http://arxiv.org/abs/1802.07589 (accessed 4 Dec2020).

51. ImageNet. http://www.image-net.org/

52. 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.

53. Psaltis D, Brady D, Gu X-G, Lin S. Holography in artificial neural networks. *Nature* 1990; **343**: 325–330.

54. Wagner KH. Deep Learning Optical Devices and Architectures. In: *Frontiers in Optics 2017 (2017), paper FW2C.1*. Optical Society of America, 2017, p FW2C.1.

55. Shastri BJ, Tait AN, de Lima TF, Nahmias MA, Peng H-T, Prucnal PR. Principles of Neuromorphic Photonics. *arXiv:180100016 [physics]* 2017. doi:10.1007/978-3-642-27737-5_702-1.

56. Hermans M, Burm M, Van Vaerenbergh T, Dambre J, Bienstman P. Trainable hardware for dynamical computing using error backpropagation through physical media. *Nature Communications* 2015; **6**: 6729.

57. Brunner D, Soriano MC, Mirasso CR, Fischer I. Parallel photonic information processing at gigabyte per second data rates using transient states. *Nature Communications* 2013; **4**: 1364.

58. 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. *Opt Mater Express, OME* 2012; **2**: 942–947.

59. Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. *arXiv:14126980 [cs]* 2017.http://arxiv.org/abs/1412.6980 (accessed 4 Dec2020).

60. Kazhdan M, Hoppe H. Screened poisson surface reconstruction. *ACM Trans Graph* 2013; **32**: 29:1–29:13.

61. Cignoni P, Callieri M, Corsini M, Dellepiane M, Ganovelli F, Ranzuglia G. MeshLab: an Open-Source Mesh Processing Tool. *Eurographics Italian Chapter Conference* 2008; : 8 pages.

62. Goodman J. Introduction to Fourier optics. 1968 doi:10.1063/1.3035549.

63. Yu N, Capasso F. Flat optics with designer metasurfaces. *Nature Materials* 2014; **13**: 139–150.

64. Khorasaninejad M, Chen WT, Devlin RC, Oh J, Zhu AY, Capasso F. Metalenses at visible wavelengths: Diffraction-limited focusing and subwavelength resolution imaging. *Science* 2016; **352**: 1190–1194.

65. Kildishev AV, Boltasseva A, Shalaev VM. Planar Photonics with Metasurfaces. *Science* 2013; **339**. doi:10.1126/science.1232009.

66. Trabelsi C, Bilaniuk O, Zhang Y, Serdyuk D, Subramanian S, Santos JF *et al.* Deep Complex Networks. *arXiv:170509792 [cs]* 2018.http://arxiv.org/abs/1705.09792 (accessed 4 Dec2020).

67. Zhou Wang, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing* 2004; **13**: 600–612.

68. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res* 2014; **15**: 1929–1958.

69. Grischkowsky D, Keiding S, Exter M van, Fattinger C. Far-infrared time-domain spectroscopy with terahertz beams of dielectrics and semiconductors. *J Opt Soc Am B, JOSAB* 1990; **7**: 2006–2015.

70. Bianchi V, Carey T, Viti L, Li L, Linfield EH, Davies AG *et al.* Terahertz saturable absorbers from liquid phase exfoliation of graphite. *Nature Communications* 2017; **8**: 15763.

71. Marini A, Cox JD, García de Abajo FJ. Theory of graphene saturable absorption. *Phys Rev B* 2017; **95**: 125408.

72. Yin X, Feng T, Liang Z, Li J. Artificial Kerr-type medium using metamaterials. *Opt Express, OE* 2012; **20**: 8543–8550.

73. Xiao Y, Qian H, Liu Z. Nonlinear Metasurface Based on Giant Optical Kerr Response of Gold Quantum Wells. *ACS Photonics* 2018; **5**: 1654–1659.

74. Cox DD, Dean T. Neural Networks and Neuroscience-Inspired Computer Vision. *Current Biology* 2014; **24**: R921–R929.

75. 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*. Association for Computing Machinery: Helsinki, Finland, 2008, pp 160–167.

76. 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.

77. Nehme E, Weiss LE, Michaeli T, Shechtman Y. Deep-STORM: super-resolution single-molecule microscopy by deep learning. *Optica, OPTICA* 2018; **5**: 458–464.

78. Ouyang W, Aristov A, Lelek M, Hao X, Zimmer C. Deep learning massively accelerates super-resolution localization microscopy. *Nature Biotechnology* 2018; **36**: 460–468.

79. Wang H, Rivenson Y, Jin Y, Wei Z, Gao R, Günaydın H *et al.* Deep learning enables cross-modality super-resolution in fluorescence microscopy. *Nature Methods* 2019; **16**: 103–110.

80. 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.

81. 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: Science & Applications* 2019; **8**: 23.

82. Wu Y, Ray A, Wei Q, Feizi A, Tong X, Chen E *et al.* Deep Learning Enables High-Throughput Analysis of Particle-Aggregation-Based Biosensors Imaged Using Holography. *ACS Photonics* 2019; **6**: 294–301.

83. 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**: 25.

84. 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, OPTICA* 2018; **5**: 704–710.

85. Ballard ZS, Joung H-A, Goncharov A, Liang J, Nugroho K, Carlo DD *et al.* Deep learning-enabled point-of-care sensing using multiplexed paper-based sensors. *npj Digital Medicine* 2020; **3**: 1–8.

86. Holmström O, Linder N, Ngasala B, Mårtensson A, Linder E, Lundin M *et al.* Point-of-care mobile digital microscopy and deep learning for the detection of soil-transmitted helminths and Schistosoma haematobium. *Global Health Action* 2017; **10**: 1337325.

87. 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.

88. Veli M, Ozcan A. Computational Sensing of Staphylococcus aureus on Contact Lenses Using 3D Imaging of Curved Surfaces and Machine Learning. *ACS Nano* 2018; **12**: 2554–2559.

89. 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.

90. Peurifoy J, Shen Y, Jing L, Yang Y, Cano-Renteria F, DeLacy BG *et al.* Nanophotonic particle simulation and inverse design using artificial neural networks. *Science Advances* 2018; **4**: eaar4206.

91. 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.

92. 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.

93. Borhani N, Kakkava E, Moser C, Psaltis D. Learning to see through multimode fibers. *Optica, OPTICA* 2018; **5**: 960–966.

94. Hughes TW, Williamson IAD, Minkov M, Fan S. Wave physics as an analog recurrent neural network. *Science Advances* 2019; **5**: eaay6946.

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]. *Opt Mater Express, OME* 2018; **8**: 3851–3863.

96. 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.

97. Sande GV der, Brunner D, Soriano MC. Advances in photonic reservoir computing. *Nanophotonics* 2017; **6**: 561–576.

98. Hamerly R, Bernstein L, Sludds A, Soljačić M, Englund D. Large-Scale Optical Neural Networks Based on Photoelectric Multiplication. *Phys Rev X* 2019; **9**: 021032.

99. 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.

100. Li J, Mengu D, Luo Y, Rivenson Y, Ozcan A. Class-specific differential detection in diffractive optical neural networks improves inference accuracy. *Adv Photonics* 2019; **1**: 046001.

101. 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* 2020; **26**: 1–14.

102. Marin-Palomo P, Kemal JN, Karpov M, Kordts A, Pfeifle J, Pfeiffer MHP *et al.* Microresonator-based solitons for massively parallel coherent optical communications. *Nature* 2017; **546**: 274–279.

103. Strickland D, Mourou G. Compression of amplified chirped optical pulses. *Optics Communications* 1985; **56**: 219–221.

104. Vabishchevich PP, Shcherbakov MR, Bessonov VO, Dolgova TV, Fedyanin AA. Femtosecond pulse shaping with plasmonic crystals. *Jetp Lett* 2015; **101**: 787–792.

105. Rahimi E, Şendur K. Femtosecond pulse shaping by ultrathin plasmonic metasurfaces. *J Opt Soc Am B, JOSAB* 2016; **33**: A1–A7.

106. Szipöcs R, Ferencz K, Spielmann C, Krausz F. Chirped multilayer coatings for broadband dispersion control in femtosecond lasers. *Opt Lett, OL* 1994; **19**: 201–203.

107. Supradeepa VR, Huang C-B, Leaird DE, Weiner AM. Femtosecond pulse shaping in two dimensions: Towards higher complexity optical waveforms. *Opt Express, OE* 2008; **16**: 11878–11887.

108. Weiner AM. Femtosecond pulse shaping using spatial light modulators. *Review of Scientific Instruments* 2000; **71**: 1929–1960.

109. Dugan MA, Tull JX, Warren WS. High-resolution acousto-optic shaping of unamplified and amplified femtosecond laser pulses. *J Opt Soc Am B, JOSAB* 1997; **14**: 2348–2358.

110. Weiner AM. Ultrafast optical pulse shaping: A tutorial review. *Optics Communications* 2011; **284**: 3669–3692.

111. Yelin D, Meshulach D, Silberberg Y. Adaptive femtosecond pulse compression. *Opt Lett, OL* 1997; **22**: 1793–1795.

112. Assion A, Baumert T, Bergt M, Brixner T, Kiefer B, Seyfried V *et al.* Control of Chemical Reactions by Feedback-Optimized Phase-Shaped Femtosecond Laser Pulses. *Science* 1998; **282**: 919–922.

113. Efimov A, Moores MD, Beach NM, Krause JL, Reitze DH. Adaptive control of pulse phase in a chirped-pulse amplifier. *Opt Lett, OL* 1998; **23**: 1915–1917.

114. Weiner AM, Leaird DE, Patel JS, Wullert JR. Programmable femtosecond pulse shaping by use of a multielement liquid-crystal phase modulator. *Opt Lett, OL* 1990; **15**: 326–328.

115. Bardeen CJ, Yakovlev VV, Wilson KR, Carpenter SD, Weber PM, Warren WS. Feedback quantum control of molecular electronic population transfer. *Chemical Physics Letters* 1997; **280**: 151–158.

116. Hillegas CW, Tull JX, Goswami D, Strickland D, Warren WS. Femtosecond laser pulse shaping by use of microsecond radio-frequency pulses. *Opt Lett, OL* 1994; **19**: 737–739.

117. Zeek E, Maginnis K, Backus S, Russek U, Murnane M, Mourou G *et al.* Pulse compression by use of deformable mirrors. *Opt Lett, OL* 1999; **24**: 493–495.

118. Divitt S, Zhu W, Zhang C, Lezec HJ, Agrawal A. Ultrafast optical pulse shaping using dielectric metasurfaces. *Science* 2019; **364**: 890–894.

119. Hashemi MR, Cakmakyapan S, Jarrahi M. Reconfigurable metamaterials for terahertz wave manipulation. *Rep Prog Phys* 2017; **80**: 094501.

120 Rahm M, Li J-S, Padilla WJ. THz Wave Modulators: A Brief Review on Different Modulation Techniques. *J Infrared Milli Terahz Waves* 2013; **34**: 1–27.

121. Danielson JR, Amer N, Lee Y-S. Generation of arbitrary terahertz wave forms in fanned-out periodically poled lithium niobate. *Appl Phys Lett* 2006; **89**: 211118.

122. Stepanov AG, Hebling J, Kuhl J. Generation, tuning, and shaping of narrow-band, picosecond THz pulses by two-beam excitation. *Opt Express, OE* 2004; **12**: 4650–4658.

123. Sato M, Higuchi T, Kanda N, Konishi K, Yoshioka K, Suzuki T *et al.* Terahertz polarization pulse shaping with arbitrary field control. *Nature Photonics* 2013; **7**: 724–731.

124. Keren-Zur S, Tal M, Fleischer S, Mittleman DM, Ellenbogen T. Generation of spatiotemporally tailored terahertz wavepackets by nonlinear metasurfaces. *Nature Communications* 2019; **10**: 1778.

125. Yongqian Liu, Sang-Gyu Park, Weiner AM. Terahertz waveform synthesis via optical pulse shaping. *IEEE Journal of Selected Topics in Quantum Electronics* 1996; **2**: 709–719.

126. Gingras L, Cooke DG. Direct temporal shaping of terahertz light pulses. *Optica, OPTICA* 2017; **4**: 1416–1420.

127. Burghoff D, Kao T-Y, Han N, Chan CWI, Cai X, Yang Y *et al.* Terahertz laser frequency combs. *Nature Photonics* 2014; **8**: 462–467.

128. Bachmann D, Rösch M, Süess MJ, Beck M, Unterrainer K, Darmo J *et al.* Short pulse generation and mode control of broadband terahertz quantum cascade lasers. *Optica, OPTICA* 2016; **3**: 1087–1094.

129. Barbieri S, Ravaro M, Gellie P, Santarelli G, Manquest C, Sirtori C *et al.* Coherent sampling of active mode-locked terahertz quantum cascade lasers and frequency synthesis. *Nature Photonics* 2011; **5**: 306–313.

130. van der Weide DW. Delta-doped Schottky diode nonlinear transmission lines for 480-fs, 3.5-V transients. *Appl Phys Lett* 1994; **65**: 881–883.

131. Assefzadeh MM, Babakhani A. Broadband Oscillator-Free THz Pulse Generation and Radiation Based on Direct Digital-to-Impulse Architecture. *IEEE Journal of Solid-State Circuits* 2017; **52**: 2905–2919.

132. Abo-Bakr M, Feikes J, Holldack K, Kuske P, Peatman WB, Schade U *et al.* Brilliant, Coherent Far-Infrared (THz) Synchrotron Radiation. *Phys Rev Lett* 2003; **90**: 094801.

133. Yardimci NT, Jarrahi M. High Sensitivity Terahertz Detection through Large-Area Plasmonic Nano-Antenna Arrays. *Scientific Reports* 2017; **7**: 42667.

134. Cunningham PD, Valdes NN, Vallejo FA, Hayden LM, Polishak B, Zhou X-H *et al.* Broadband terahertz characterization of the refractive index and absorption of some important polymeric and organic electro-optic materials. *Journal of Applied Physics* 2011; **109**: 043505-043505–5.

135. Podzorov A, Gallot G. Low-loss polymers for terahertz applications. *Appl Opt, AO* 2008; **47**: 3254–3257.

136. Jin Y-S, Kim G-J, Jeon S-G. Terahertz dielectric properties of polymers. *Journal of the Korean Physical Society* 2006; **49**: 513–517.

137. Partanen A, Väyrynen J, Hassinen S, Tuovinen H, Mutanen J, Itkonen T *et al.* Fabrication of terahertz wire-grid polarizers. *Appl Opt, AO* 2012; **51**: 8360–8365.

138. Kitahara H, Tsumura N, Kondo H, Takeda MW, Haus JW, Yuan Z *et al.* Terahertz wave dispersion in two-dimensional photonic crystals. *Phys Rev B* 2001; **64**: 045202.

139. Yardimci NT, Yang S-H, Berry CW, Jarrahi M. High-Power Terahertz Generation Using Large-Area Plasmonic Photoconductive Emitters. *IEEE Transactions on Terahertz Science and Technology* 2015; **5**: 223–229.

140. Kulce O, Mengu D, Rivenson Y, Ozcan A. All-Optical Information Processing Capacity of Diffractive Surfaces. *arXiv:200712813 [physics]* 2020.http://arxiv.org/abs/2007.12813 (accessed 9 Aug2020).