

UC Berkeley

UC Berkeley Previously Published Works

Title

Wide-field, high-resolution reconstruction in computational multi-aperture miniscope using a Fourier neural network.

Permalink

<https://escholarship.org/uc/item/3r98g740>

Journal

Optica, 11(6)

ISSN

2334-2536

Authors

Yang, Qianwan

Guo, Ruipeng

Hu, Guorong

et al.

Publication Date

2024-06-20

DOI

10.1364/OPTICA.523636

Peer reviewed



Published in final edited form as:

Optica. 2024 June 20; 11(6): 860–871. doi:10.1364/OPTICA.523636.

Wide-field, high-resolution reconstruction in computational multi-aperture miniscope using a Fourier neural network

Qianwan Yang¹, Ruipeng Guo¹, Guorong Hu¹, Yujia Xue¹, Yunzhe Li^{1,2}, Lei Tian^{1,3,4,*}

¹Department of Electrical and Computer Engineering, Boston University, Boston, Massachusetts 02215, USA

²Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, California 94720, USA

³Department of Biomedical Engineering, Boston University, Boston, Massachusetts 02215, USA

⁴Neurophotonics Center, Boston University, Boston, Massachusetts 02215, USA

Abstract

Traditional fluorescence microscopy is constrained by inherent trade-offs among resolution, field of view, and system complexity. To navigate these challenges, we introduce a simple and low-cost computational multi-aperture miniature microscope, utilizing a microlens array for single-shot wide-field, high-resolution imaging. Addressing the challenges posed by extensive view multiplexing and non-local, shift-variant aberrations in this device, we present SV-FourierNet, a multi-channel Fourier neural network. SV-FourierNet facilitates high-resolution image reconstruction across the entire imaging field through its learned global receptive field. We establish a close relationship between the physical spatially varying point-spread functions and the network's learned effective receptive field. This ensures that SV-FourierNet has effectively encapsulated the spatially varying aberrations in our system and learned a physically meaningful function for image reconstruction. Training of SV-FourierNet is conducted entirely on a physics-based simulator. We showcase wide-field, high-resolution video reconstructions on colonies of freely moving *C. elegans* and imaging of a mouse brain section. Our computational multi-aperture miniature microscope, augmented with SV-FourierNet, represents a major advancement in computational microscopy and may find broad applications in biomedical research and other fields requiring compact microscopy solutions.

1. INTRODUCTION

Wide-field, high-resolution imaging plays a critical role across diverse scientific disciplines. Yet, traditional microscopy, which relies on single-objective lenses, is encumbered by a fundamental trade-off between field of view (FOV) and resolution [1]. To circumvent these constraints, array microscopes have been developed, facilitating rapid, wide-FOV image

* leitian@bu.edu .

Disclosures. The authors declare no conflicts of interest.

Supplemental document. See Supplement 1 for supporting content.

acquisition [2–4]. These systems utilize an ensemble of microscopes, each having its own lens and image sensor, to expand the FOV without sacrificing resolution. Despite these advancements, the scalability and miniaturization of array microscopes are limited by their dependence on bulky, specialized image sensor arrays, which pose significant challenges in applications demanding portability and compactness, such as endoscopy [5], on-chip microscopy [6], and *in vivo* neural imaging [7]. Recently, lensless imaging techniques have emerged as a compact alternative by employing a phase mask [8,9] placed directly in front of a CMOS sensor. However, the removal of focusing optics leads to reduced contrast and signal-to-noise ratio (SNR), which restricts their sensitivity for imaging weak fluorescent signals [10,11].

To address these challenges, we introduce a computational multi-aperture miniature microscope (miniscope) that delivers wide FOV and micrometer resolution on a compact and lightweight device. Drawing inspiration from array microscopes, our design employs a microlens array (MLA) as the sole imaging component. Leveraging principles from lightfield and integral imaging techniques [12], our previous work has demonstrated single-shot 3D imaging capability using such a design [13,14]. This work aims to provide a new perspective of this multi-aperture imaging system in order to address the challenges related to FOV and uneven resolution. Our approach is based on the premise that each lens within the MLA can capture a high-quality image of a distinct, albeit limited, sub-FOV [Fig. 1(a)]. While individual lenses yield images that suffer from spatially varying aberrations and contrast loss towards the edges of their respective sub-FOVs, combining information from various microlenses enables the reconstruction of a high-resolution image over an extended FOV, as illustrated in Figs. 1(a) and 1(c).

Departing from conventional array microscopes, our system utilizes a single image sensor to enhance device miniaturization and portability. Unlike other MLA-based multi-aperture imaging designs [15–17] that employ physical barriers to isolate sub-images captured by individual microlenses, our system opts for an unobstructed arrangement to maximize the effective imaging FOV [13]. This design, however, introduces severe cross-talk between adjacent sub-images on the sensor, i.e., view multiplexing. Furthermore, the single-element MLA generates point-spread functions (PSFs) that contain widely separated array foci and exhibit considerable spatial variability, i.e., non-local and spatially varying PSFs. These complexities pose substantial challenges to achieving high-quality image reconstruction across the extended FOV.

Our solution to simultaneously tackle view multiplexing and non-local, spatially varying aberrations is to employ a deconvolution algorithm. Traditional model-based methods rely on iterative algorithms [18–20], which are computationally intensive and time-consuming. Additionally, the reconstruction quality is greatly affected by manually tuned regularization terms.

Recent progress in deep-learning-based spatially varying deconvolution methods have shown improvements in image quality and reconstruction speed [21,22], emerging as a promising alternative to model-based methods. However, deep-learning methods often suffer from two major limitations in terms of *scalability* and *locality bias*. Specifically, the

computational complexity and memory demands of convolutional neural networks (CNNs) increase with the input size. Thus, the direct application of CNNs to raw measurements containing tens of millions of pixels is impractical due to limited computational resources. The common workaround involves dividing raw images into smaller patches to make CNN training manageable. This method, however, assumes uniformity in the data distribution, and thus *invariance* in the PSFs for all image patches, overlooking the nuances due to spatially varying aberrations. Consequently, such CNN-based methods often exhibit degraded resolution at FOV periphery, where aberrations diverge from those in central regions [14].

A second challenge in using CNNs for handling non-local, spatially varying PSFs is the networks' inherent locality bias [23,24], a result of the small filter sizes in convolution layers that restrict CNNs from capturing global information. Although increasing the CNN's depth can broaden the "receptive field," this approach escalates computational costs and complicates network training. The recently proposed MultiWienerNet and MultiFlatNet [21,22] address these issues by initially applying deconvolution with multiple Wiener filters informed by a set of pre-calibrated, spatially varying PSFs, followed by refinement through a CNN. Yet, this type of technique heavily depends on the choice of these initial PSFs that need to be sufficiently representative of the underlying spatially varying aberrations across the FOV. Our study also shows that this *spatial* domain learning approach cannot efficiently reconstruct high-resolution information due to the implicit bias towards low-frequency features induced by the Wiener filtering process. An alternative approach is the *Fourier* domain neural network (FourierNet), which overcomes locality bias by applying filters directly in the Fourier domain [25–28]. This strategy explicitly takes advantage of the *global* processing capabilities of the Fourier transform, allowing FourierNet to achieve a global receptive field with minimal layers, thus efficiently sidestepping the limitations faced by traditional CNNs. While FourierNet has been successfully applied in spatially *invariant* imaging systems [26–28], its ability to address spatially varying systems remains unexplored.

In this work, we introduce spatially varying FourierNet (SV-FourierNet) to facilitate wide-field, high-resolution reconstruction within our computational multi-aperture miniscope. We show that SV-FourierNet can be applied directly on full-FOV measurements containing 12 million pixels, circumventing the uneven resolution constraints typically associated with patch-based training methods. Furthermore, by employing multiple learnable Fourier filters [Fig. 1(b)], SV-FourierNet is effective in addressing the view multiplexing and highly non-local and spatially varying array PSFs in our multi-aperture miniscope. The multi-channel Fourier layer in SV-FourierNet effectively performs a blind deconvolution, based on a low-rank imaging model, to process the input measurement. The resulting multi-channel deconvolved outputs are then fused and refined by an enhancement network, yielding the final reconstructed image. Our study shows that this *Fourier* domain learning approach can more efficiently capture high-resolution information as compared to the spatial domain learning methods.

To elucidate the working principle of SV-FourierNet, we employ a gradient-based method to probe the network's "effective receptive field (ERF)" [29]. This method identifies the points

in the measurement that most significantly affect the prediction of a specific point on the reconstructed object, closely resembling the PSF of our imaging system at the corresponding location. We show a close correspondence between SV-FourierNet's ERFs and the physical PSFs at various field positions. This underscores that SV-FourierNet has encapsulated the spatially varying aberrations in our system and learned a *physically meaningful* function for image reconstruction.

Compared to current existing state-of-the-art model-based and deep-learning-based methods [14,21,22], SV-FourierNet demonstrates superior reconstruction quality, consistently higher resolution across the FOV, and enhanced inference speed. Trained entirely on a physics-based simulator, we show experimentally that SV-FourierNet attains a uniform resolution of 7.8 μm across a 6.5 mm FOV. The network also provides consistent reconstruction quality across an extended depth of field (EDOF) over a 100 μm depth range. Additionally, we present wide-field, high-resolution video reconstructions of freely moving *C. elegans* colonies and imaging of a weakly scattering mouse brain section.

Our computational multi-aperture miniature microscope augmented with SV-FourierNet represents a significant advancement in computational microscopy, promising wide-ranging applications in biomedical research and other areas requiring compact imaging solutions.

2. METHODS

A. Multi-Aperture Miniscope

The multi-aperture miniscope is built with off-the-shelf and 3D-printed components. It uses an off-the-shelf plastic 3×3 MLA (1 mm pitch, No. 630, Fresnel Technologies Inc.), forming a finite conjugate imaging geometry with a working distance of ~ 7.5 mm and an MLA-sensor distance of ~ 4.5 mm. Each lens within the MLA captures an image of a distinct and limited sub-FOV, while the whole array covers an extended FOV (see more details in Section 1 in Supplement 1). For green fluorescence imaging, a hybrid filter set is incorporated, combining an interference filter (No. 535/50, Chroma Technology) placed in front of the MLA and a long-pass absorption filter (Wratten color filter No. 12, Edmund Optics) placed after the MLA. It uses a backside-illuminated (BSI) CMOS sensor (IM226, IDS Imaging) with 12 million pixels and 1.85 μm pixel size, achieving an effective pixel size of 2.8 μm . To validate the reconstruction results from our system, we set up a wide-field microscope to record fluorescence simultaneously with our miniscope. See more details in Section 9A–B in Supplement 1.

B. Spatially Varying Imaging Model

To bypass the need for extensive experimental data collection, we develop a highly efficient and accurate simulator to generate a comprehensive training dataset *in silico*. Building upon our previous work [14], this simulator incorporates a low-rank spatially varying imaging model, improved with insights from a ray-tracing model, as detailed below.

We first collect a set of calibrated PSFs by scanning a single fluorescent bead across a 7 mm \times 7 mm FOV with a uniform step size of 250 μm . To showcase the strong spatial variance of the PSFs, we calculate the Pearson coefficient correlation (PCC) map on the calibrated PSFs

with respect to the on axis PSF, as shown in Fig. 2(a). The PCC map exhibits a rapid radial decay due to the limited angular response of the system and spatially varying aberrations from the MLA. The central line profile shows that the PCC decreases to 0.2 with a 3.5 mm displacement.

Next, we develop a low-rank model in three steps. To enhance computational efficiency, we first exploit the sparsity in the array PSF generated by the MLA and perform a cropping operation based on a ray tracing model to isolate each focal region (240×240 pixels) from the raw PSF measurement [Fig. 2(b)]. This method relies on the calibration of the chief ray of each microlens (ML) (i, j) from an on-axis point source, which establishes the “anchor” coordinates in both the image space ($p_{x_0}^{i,j}, q_{y_0}^{i,j}$) and the object space (x_0, y_0). Given the depth z and the point source’s lateral coordinates (x, y), the corresponding PSF centroids for each ML (i, j)($p_x^{i,j}, q_y^{i,j}$) are linearly proportional to the magnification M_z and the object displacement:

$$\begin{aligned} p_x^{i,j} &= p_{x_0}^{i,j} - M_z \cdot (x - x_0), \\ q_y^{i,j} &= q_{y_0}^{i,j} - M_z \cdot (y - y_0). \end{aligned} \quad (1)$$

Contrasting with cross-correlation methods typically employed to locate centroids of calibrated PSFs [14,30], our ray tracing model embeds the image distortion information into the PSF patches. The distortion is characterized by the subsequent decomposition process and corrected during the reconstruction phase.

Second, to approximate the system’s spatial variance, we implement truncated singular value decomposition (TSVD) on the cropped foci for each individual ML, denoted as $H^{i,j}(p, q; x, y)$. Unlike our previous work [14] that performs a global decomposition on the array PSFs, our new method decomposes the foci of each ML independently. This is based on the observation that aberrations tend to be more consistent within a single ML but vary across MLs, largely due to the manufacturing variations, as shown in Fig. 2(b). Exploiting these distinct aberrations, our single-ML decomposition method effectively reduces the rank of the imaging model and computational complexity by $2\times$ without compromising the model accuracy [Fig. 2(c)]. The decomposition process can be written as

$$H^{i,j}(p, q; x, y) \approx \sum_{k=1}^K M_k^{i,j}(x, y) B_k^{i,j}(p, q), \quad (2)$$

where $M_k^{i,j}(x, y)$ and $B_k^{i,j}(p, q)$ denote the k th coefficient map and basis PSF for ML (i, j), respectively. We select $K = 32$, achieving an approximation error of $\sim 2.5\%$ on the calibration set.

Finally, the spatially varying model is implemented in two steps, as shown in Fig. 2(d). First, the object is element-wise multiplied with a coefficient map and then convolved with the

corresponding basis PSF for each ML. The K weighted convolutions are summed to form a single-ML measurement. Next, all the single ML measurements are “placed back” according to the pre-calibrated anchor location $(p_{s_0}^{i,j}, q_{s_0}^{i,j})$, and then summed together to construct the final multiplexed measurement. This process is described as

$$g(p, q) = \sum_{i,j=1}^3 \sum_{k=1}^K [M_k^{i,j}(p, q) \odot O(p, q)] \oplus_{p,q} B_k^{i,j}(p, q), \quad (3)$$

where $O(p, q)$ is the object’s fluorescence distribution, and the image space coordinates (p, q) are related to the object space coordinates (x, y) by Eq. (1).

To expedite the simulation process, sub-images from individual MLs are computed in parallel, allowing the model to generate a measurement within 30 s. The sensor noise is sampled from a pre-calibrated mixed Poisson–Gaussian noise model. The noise is added to the measurements in the data loader as a form of “online” data augmentation during the training process, which ensures that the synthetic measurements closely approximate real imaging conditions (see details in Section 3.B in Supplement 1).

The spatially varying imaging model is validated by comparing the simulated and experimental measurements on the same object containing randomly distributed particles, shown in Fig. 2(e) (see details in Section 3.A in Supplement 1). The agreement between simulation and experiment, measured by PCC, achieves 0.92, indicating our model’s accuracy. Furthermore, two zoomed-in regions show that the model accurately reproduces the aberrations present in the physical measurement.

To promote generalization of SV-FourierNet, we simulate training data by applying our low-rank model to a diverse set of biological images collected from online repositories [31–35], including cells, vasculature, and brain sections, as well as synthetically generated fluorescent particles with random size, brightness, and locations, yielding around 5000 training images (see details in Section 3.C in Supplement 1).

C. SV-FourierNet Design

SV-FourierNet features high scalability and effective processing of global spatial information, enabling the direct processing of raw measurements containing 12 million pixels. This capability allows it to simultaneously address the challenges of view multiplexing and non-local, spatially varying aberrations in our system.

SV-FourierNet comprises two core modules, including a blind Fourier deconvolution module and an enhancement module, as shown in Fig. 3. The measurement is first Fourier transformed into the frequency domain, followed by element-wise multiplication with multiple learnable complex Fourier filters, whose weights are initialized with a uniform distribution $U[0,1]$. The filtered results are then inverse Fourier transformed back to the spatial domain, resulting in the initial multi-channel reconstructions.

The Fourier-domain blind deconvolution approach in SV-FourierNet, in contrast to MultiWienerNet/MultiFlatNet [21,22], bypasses the need for pre-selecting PSFs and facilitates the efficient learning of a low-rank Fourier basis. Additionally, by leveraging the conjugate symmetry of the Fourier transform, SV-FourierNet halves the number of training parameters required compared to those needed for learning directly in the spatial domain [21,22]. SV-FourierNet employs nine Fourier filters to collectively account for the spatial variance in our system. The number of filters is crucial as it dictates the network's capacity for nonlinearity. We have conducted an ablation study on the number of Fourier filters, which revealed that increasing the number of filters improves resolution and reconstruction quality but at the cost of memory consumption. Given the trade-off between network performance and computational cost, we have opted for nine channels to facilitate the spatially varying reconstruction (see details in Section 4.C in Supplement 1).

The enhancement module employs a modified residual channel attention network (RCAN) [36], which includes a down-scaling module, a residual in residual (RIR) block, and an up-scaling module. Each RIR block connects two residual groups with a long skip connection, facilitating information fusion across different scales. Channel attention modules embedded within the residual groups dynamically adjust the channel-wise feature weights, essential for synthesizing information from different Fourier channels. To prevent overfitting and to improve generalizability, the enhancement module is designed to be lightweight, with approximately 0.18 million parameters. Additional details about the network implementation are in Section 4.B in Supplement 1. In addition, cropping and padding operations are implemented in the data loader to address the periodic ambiguity characteristic of array PSFs (see details in Section 4.A in Supplement 1).

The loss function is a sum of binary cross entropy (BCE) and mean squared error (MSE):

$$L_{\text{total}} = L_{\text{BCE}} + L_{\text{MSE}}, \quad (4)$$

$$L_{\text{BCE}} = \frac{1}{n} \sum_{i=1}^n [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)], \quad (5)$$

$$L_{\text{MSE}} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (6)$$

where y_i and \hat{y}_i represent the true and predicted values, respectively, i indexes the pixel, and n is the total number of pixels. This dual-loss approach leverages BCE for enhancing the reconstruction of sharp features and MSE for capturing intensity variations, guiding the model towards accurate reconstructions in both structure and intensity. Gradient

backpropagation from this loss updates the weights of both the Fourier filters and the enhancement module in an end-to-end manner.

We implement SV-FourierNet in PyTorch on an Nvidia A40 GPU. Training is conducted entirely on a simulated dataset with a batch size of 4. Optimization employs the Adam optimizer alongside a cosine annealing scheduler, with the entire training process spanning approximately 48 h. (See detailed loss curve in Section 4.D in Supplement 1).

D. Computation of the ERF

To assess SV-FourierNet’s ability to encapsulate non-local, spatially varying aberrations and learn a physically meaningful inverse mapping, we calculate its ERF after the network is trained. This involves gradient backpropagation from a pixel on the reconstructed object to the input measurement [29], effectively measuring the influence of the measurement on a specific object pixel and generating a localized heatmap. In practice, this heatmap not only depends on the network’s trained weights, but also varies with different input objects, making it input-dependent. The ERF is the ensemble average of the heatmaps across a large dataset, effectively making it input-*independent*. The SV-FourierNet’s ERF is computed using a set of 520 test images. (See detailed implementation in Section 5.A in Supplement 1).

We emphasize that the ERF-based method of visualizing the network’s response offers more insights into the network’s learned functional mapping compared to other techniques such as the saliency map [37] and Grad-CAM [38], which often produce more ambiguous heatmaps and are designed to operate on an individual input basis. The strength of our ERF-based analysis lies in its capacity to calculate a statistically averaged response, making the resulting heatmaps agnostic to the input. If the network fails to learn a meaningful mapping, the most likely ERF is a feature-less blur across a substantial portion of the input field, a consequence of the averaging process [29]. The highly localized ERFs suggests that SV-FourierNet has learned to generalize across different objects, and the learned mapping function aligns closely with the underlying imaging model.

E. Model-Based Reconstruction Algorithms

We derive and implement an LSV model-based algorithm based on solving a regularized least-squares problem using the alternating direction method of multiplier (ADMM) algorithm, as detailed in Section 2 in Supplement 1.

3. RESULTS

A. SV-FourierNet’s Learned ERF Aligns with Physical PSFs

We observe that the ERFs of the network closely resemble the definition of the PSFs of the physical imaging system. To discern whether SV-FourierNet has learned a meaningful reconstruction function beyond mere image feature recognition, we directly compare SV-FourierNet’s ERFs with the system’s PSFs across different field locations.

SV-FourierNet’s ERFs contain both positive and negative components. The positive ERF components indicate that specific pixels in the measurement lead to an increase in intensity

at the reconstructed pixel by SV-FourierNet. Given that PSFs in fluorescence imaging are inherently positive, it is the positive ERF components that parallel the physical PSFs function. As shown in Fig. 4, SV-FourierNet's ERFs are highly localized and closely match the physically measured PSFs across the field. At the FOV center, the ERF aligns with the physical PSF, containing a widely distributed 3×3 foci array, as shown in Fig. 4(a). This demonstrates that SV-FourierNet has learned a global receptive field to synthesize information across the entire measurement and the network's effective response agrees with the physical response of the imaging system. Near the FOV edges, due to the sensor's finite size and the system's limited angular response, the physical PSFs are truncated to 2×3 or 3×2 foci array, respectively, as shown in Figs. 4(b) and 4(c). Nevertheless, the ERFs precisely localize the physical foci. This showcases SV-FourierNet's capability to learn the severe spatial variance in the imaging model, including handling image truncation—a case that often induces numerical artifacts in model-based methods [39]. In addition, when assessing a point outside the SV-FourierNet's trained FOV, the physical PSF appears as a truncated 2×2 foci array with noticeable contrast loss [Fig. 4(d)]. However, the ERF aligns precisely with this pattern, underscoring SV-FourierNet's ability to learn a physically accurate inverse mapping. This ensures model generalization even beyond the trained FOV boundaries. Zoomed-in views of the foci regions further elucidate the severe aberrations and spatial variability of the PSFs, along with the precise co-localization between the ERFs and the physical PSFs. A detailed quantitative analysis and the evolution of ERF throughout the training process are provided in Section 5B–C in Supplement 1.

The negative ERF components demonstrate SV-FourierNet's capability to suppress aberrations and demultiplex overlapping views in the measurement. These negative components, as detailed in Section 5.D in Supplement 1, either surround the positive ERF to enhance resolution or create an additional periodic array foci pattern that corresponds to the MLA's periodicity. This pattern arises from the measurement's ambiguity caused by the MLA's periodic structure, where displacement by a distance equal to the foci array's period results in nearly identical measurements. The differences are only due to unique aberrations of the individual microlenses and the image truncation by the system. SV-FourierNet navigates this complexity by leveraging distinct aberration patterns across microlenses, accentuating the accurate object components through positive ERF components while suppressing contributions from ambiguous components via negative ERF components.

Finally, to elucidate the physical mapping function learned by the blind deconvolution module of SV-FourierNet, excluding the nonlinearity induced by the enhancement network module, we compute and visualize the equivalent spatial domain filters by transforming the learned Fourier filters to the real space, as detailed in Section 6.A of Supplement 1. Notably, the visualization reveals that almost all the spatial domain filters exhibit 3×5 foci, which contrasts with the physical PSFs that contain 3×3 (or fewer) foci. To further understand this result, we compute the pseudo-inverse of the first nine basis PSFs derived via TSVD from experimentally calibrated PSFs, as detailed in Section 6.B of Supplement 1. These basis filters from our physical model show a high degree of similarity to the filters learned by SV-FourierNet, each also displaying a 3×5 array of foci. Thus, our physical interpretation of the learned Fourier filters is a low-rank approximation of the pseudo-inverse of the LSV imaging model.

B. Demonstration of Consistent Resolution Across a Wide FOV and an EDOF

To showcase our system's capability to achieve uniform resolution across a wide FOV, we characterize the resolution at both the FOV center and edges by imaging a fluorescence resolution target, as shown in Fig. 5(a). We also compare the reconstructions with the raw measurements from the central lens, which underscores the resolution enhancement and FOV expansion enabled by SV-FourierNet. In the central FOV, SV-FourierNet's reconstruction surpasses those from the central lens in both resolution and contrast, illustrating the network's efficiency in effectively synthesizing information from multiple lenses for high-resolution reconstructions. At the FOV periphery, where the central lens exhibits pronounced aberration and vignetting artifacts, SV-FourierNet robustly reconstructs the target with high resolution, showcasing the network's capability to account for strong spatially varying aberrations. The line profiles further affirm SV-FourierNet's capacity to achieve a consistent $7.8 \mu\text{m}$ resolution vertically and horizontally at both central and peripheral FOV regions. (See resolution characterization at additional FOV positions in Section 8.C in Supplement 1).

In addition, we characterize the reconstructed resolution of SV-FourierNet by computing the power spectral density (PSD). This analysis is performed on patches at different radial displacements and across various experimental samples. The PSD exhibits minimal variation, indicating that SV-FourierNet achieves consistent resolution across the FOV and demonstrates robustness to different experimental samples. (See details in Section 8.D of Supplement 1).

Furthermore, we show that the system achieves consistent resolution across an EDOF. We characterize the system resolution across the depth by axially scanning the fluorescent resolution target with $10 \mu\text{m}$ step size, as shown in Fig. 5(b). SV-FourierNet maintains consistent resolution within the $[-50 \mu\text{m}, 50 \mu\text{m}]$ depth range, as indicated in the zoomed-in regions labeled by a white box. Beyond this depth range, the resolution begins to gradually degrade, and the background exhibits ghosting artifacts, as marked by the red arrow. (See additional resolution characterization across an EDOF on FOV edges in Section 8.B in Supplement 1). These results demonstrate that SV-FourierNet, although trained at a single depth, can effectively generalize to an EDOF, providing robustness against depth variations and tilt during experiments.

C. SV-FourierNet Achieves State-of-the-Art Reconstruction Quality and Speed

To demonstrate the superior performance of SV-FourierNet, we compare the reconstructions from our network with the LSV-model-based ADMM algorithm and the existing state-of-the-art reconstruction networks, including CM²Net [14] and MultiWienerNet [21]. We demonstrate that SV-FourierNet achieves the best reconstruction quality with the fastest inference speed on diverse types of samples in simulation and experiments.

To ensure a fair comparison between the deep learning models, we employ the same loss function and training strategy across all models. Each model is first trained on the same dataset for 48 h. The optimal models achieving the highest SSIM are then chosen for the benchmark comparison (see details in Section 7 in Supplement 1). Moreover, the

architecture of the enhancement network within MultiWienerNet is modified to be the same as that of SV-FourierNet. The detailed architecture of the benchmark networks can be found in Section 7 in Supplement 1.

We compare the reconstruction results from CM²Net, MultiWienerNet, SV-FourierNet and the LSV ADMM algorithm on both simulated testing data and experimental measurements, shown in Fig. 6. From metrics evaluation, SV-FourierNet achieves the best PSNR and image quality. (Additional quantitative analysis on simulated testing data is provided in Section 7 in Supplement 1). From the resolution target reconstruction results, SV-FourierNet demonstrates superior resolution compared to CM²Net, and MultiWienerNet and is comparable to the LSV ADMM algorithm. Notably, at the edge of the FOV, SV-FourierNet maintains consistently high resolution, whereas other techniques exhibit degraded resolution, as detailed in Section 8.A of Supplement 1. By visual inspection, CM²Net suffers from severe ghosting artifacts and a degradation in performance at the FOV's peripheral regions. This is attributed to CM²Net's patch-based training strategy, which neglects the strong shift variance between patches. Similar artifacts are observed in reconstructions from the model-based deconvolution algorithm using a spatially invariant model (see details in Section 2.A in Supplement 1), demonstrating the importance of accounting for spatially varying degradation in the reconstruction process.

MultiWienerNet improves reconstruction performance at the FOV edge through multiple initial Wiener deconvolutions. However, it fails to preserve high-resolution details and suffers from pronounced background artifacts compared to SV-FourierNet. In addition to superior image quality, SV-FourierNet outperforms other methods with its enhanced reconstruction speed and less computational load. By leveraging the fast Fourier transform's computational efficiency, SV-FourierNet achieves a 20× speedup compared to CM²Net. Furthermore, by utilizing the Fourier transform's symmetry property for real-valued signals and avoiding the additional computational costs associated with Wiener deconvolution, SV-FourierNet reduces 2× memory consumption and achieves a 2× speedup over MultiWienerNet. (See additional comparisons for experimental and simulation results in Section 7 in Supplement 1).

To understand the superior performance of SV-FourierNet and our training strategy, we conducted extensive ablation studies, leading to two main insights: the impact of PSF/filter initialization and the choice of learning domain (spatial versus Fourier). First, initializing with physical PSFs, regardless of the learning domain, results in reduced resolution and pronounced noisy artifacts (see Section 6.D in Supplement 1). This is because physical PSF initialization introduces unwanted locality biases and inherent noise in the PSF measurements. Inspecting the learned filters in MultiWienerNet initialized with physical PSFs (Section 6.C in Supplement 1) reveals that the learning process primarily adds non-interpretable noisy features to the filters, similar to those reported in [21]. This argument is further corroborated by the ERF of MultiWienerNet (Section 5.B in Supplement 1), which shows severe noisy artifacts in the learned functions. Similar artifacts are observed when SV-FourierNet is trained with the same physical PSF initialization, instead of the random initialization used in our method. Replacing the physical PSF initialization with the same random initialization allows MultiWienerNet to converge to more physically meaningful

filters, similar to those learned by SV-FourierNet, but with reduced resolution. These results lead to our second insight: learning in the spatial domain generally results in inferior resolution compared to learning in the Fourier domain. This can be explained by two factors: the smooth PSF spectrum improves the conditioning of the learning problem, and Fourier domain learning eliminates an unnecessary Fourier transform operation in the spatial domain approach, reducing noise and error propagation.

D. Experimental Demonstration on a Colony of Freely Moving *C. elegans*

To showcase our system's capability in imaging large-scale dynamic biological processes, we image colonies of freely moving *C. elegans*. Details about the sample preparation are provided in Section 9.D in Supplement 1. The SV-FourierNet reconstruction [Fig. 7(a)] and concurrent wide-field measurements [Fig. 7(b)] are compared.

We compare the full-FOV reconstruction of a single frame from SV-FourierNet with the corresponding wide-field measurement. The reconstruction accurately recovers all the *C. elegans*, including those in central and edge FOV regions, showcasing consistent performance across the extensive FOV. To further demonstrate that our system is robust to dynamic movements and varying local contrast, we compare sequences of SV-FourierNet reconstructions and wide-field measurements for the area marked by the white box. These comparisons demonstrate that SV-FourierNet is capable of accurately recovering not only the larger, high-contrast *C. elegans* but also the younger specimens with faint fluorescence signals, as indicated by the yellow dashed box. Additionally, our system effectively recovers the dynamic movements of the *C. elegans* population, with a video shown in Visualization 1.

E. Experimental Demonstration on Fixed Brain Section

To demonstrate our system's capability for high-quality imaging of complex biological tissues, we image a 75 μm thick, weakly scattering brain section containing GFP-expressing neurons in the bed nucleus of the stria terminalis (BNST). The BNST, crucial in regulating emotional and stress responses, is bilaterally located on either side of the brain, approximately 2.5 mm apart, with each region extending about 2 mm. These regions are characterized by their intricate spatial structure and varying neuron density levels. (Detailed sample information is in Section 9.C in Supplement 1).

First, we evaluate the full-FOV SV-FourierNet reconstruction by comparing it with a 2 \times wide-field measurement. Upon visual inspection, SV-FourierNet reliably reconstructs the bilateral structure of the labeled brain regions, as shown in Fig. 8(a). We then compare the reconstructions of four zoomed-in regions from both sides of the brain section with 10 \times wide-field measurements [shown in Fig. 8(b)]. SV-FourierNet accurately reconstructs neuronal structures from both densely and sparsely populated areas. (An additional overlay comparison between SV-FourierNet reconstruction and 2 \times wide-field measurement is provided in Section 9.E of Supplement 1). These results demonstrate that SV-FourierNet, trained entirely on simulated data, can generalize to complex brain tissues, provide cellular resolutions across the entire brain section, and is robust against weak tissue scattering.

4. CONCLUSION

We present a computational multi-aperture miniscope augmented with SV-FourierNet that performs single-shot high-resolution wide-FOV imaging. Our system is simple and compact, using only a single MLA for imaging. Unlike studies focusing on exploiting the 3D imaging capability with an MLA [13], we introduce a novel computational framework to expand the FOV while ensuring uniform high resolution throughout. Our main innovation is SV-FourierNet, which employs multiple learnable Fourier filters to synthesize information over an extended FOV, addressing spatially varying aberrations and view multiplexing inherent to our system. Furthermore, we demonstrate a novel utility of the network's ERF in elucidating the mapping function it learns. We show that SV-FourierNet learns a physically meaningful reconstruction function for the spatially varying imaging model, achieving uniform resolution throughout the extended FOV. Trained exclusively on simulated data, SV-FourierNet exhibits robust generalization to diverse experimental samples, as demonstrated on resolution targets, fluorescent beads, live *C. elegans*, and brain tissue sections. This underscores SV-FourierNet's potential for a wide array of applications within biomedical research and beyond.

We show SV-FourierNet's superior capability to demultiplex cross-talks between regular MLAs and tackle severe spatially varying degradation in 2D imaging scenarios. This principle also holds promise for expanding the FOV in 3D imaging applications. For example, the Fourier lightfield microscope employs a field stop to prevent view multiplexing between lenses, which inevitably reduces the FOV in proportion to the lens count, thus limiting its imaging capacity for large-scale biological processes [40]. Our strategy suggests the possibility of removing the field stop and harnessing view multiplexing to increase the FOV without compromising spatial resolution. In 3D multi-aperture microscopes with view multiplexing, such as CM², existing 3D reconstruction algorithms suffer from resolution degradation near the FOV edges [14]. Extending SV-FourierNet to 3D imaging may substantially improve the imaging volume with uniform 3D resolution. Thus, we anticipate that SV-FourierNet could offer a novel paradigm for single-shot wide-FOV high-resolution volumetric imaging.

SV-FourierNet effectively learns global shift-variant inverse filters in the frequency domain, which not only significantly enhances image quality with higher spatial resolution but also substantially reduces the computational burden compared to spatial domain learning. This reduction is crucial for model scalability, especially when handling large volumes of data. To further decrease model size and computational load, a promising direction involves exploring the sparse nature of the system's PSFs. Such reduction in the model size will enable SV-FourierNet to be adapted to more challenging applications, such as volumetric shift-variant deconvolution.

For the first time to our knowledge, we establish a connection between the ERF of a deconvolution network and the underlying physical imaging model. This enables a straightforward visualization of the function learned by the network, alleviating the "black box" aspect often associated with the deep learning networks. Furthermore, the ERF can offer a promising tool to inform network design and allow for assessing network

performance and alignment between the trained deconvolution model and the physical optical system by visualizing deviations of the ERFs from the physical PSFs.

Our accurate and efficient physics simulator eliminates the need for extensive and laborious physical data collection, facilitating a broad data distribution, which is particularly impactful when ground truth is difficult to obtain. These simulated training data are crucial for the network's ability to generalize robustly across different experimental data types. To broaden the platform's application to more complex scattering biological tissues, it is possible to integrate the scattering process into the imaging model [41]. With training data derived from this enhanced imaging model, we anticipate the platform's utility to more challenging biomedical imaging applications.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgment.

Funded by the National Institutes of Health (R01NS126596) and a grant from 5022 - Chan Zuckerberg Initiative DAF, an advised fund of Silicon Valley Community Foundation. The authors thank Boston University Shared Computing Cluster for providing the computational resources, Ian G. Davison, Kevin J. Monk, and Brett T. DiBenedictis for providing the brain slices, as well as Christopher A. Gabel and Andrew S. Chang for providing the *C. elegans* sample. We also thank BU CISL group members for insightful discussions and Adelina Chau and Jamin Xie for their assistance with benchmark network tuning.

Funding.

National Institutes of Health (R01NS126596); Chan Zuckerberg Initiative (5022).

Data availability.

The SV-FourierNet implementation and the pretrained model are available at Ref. [42].

REFERENCES

1. Park J, Brady DJ, Zheng G, et al. , "Review of bio-optical imaging systems with a high space-bandwidth product," Adv. Photon 3, 044001 (2021).
2. Harfouche M, Kim K, Zhou KC, et al. , "Imaging across multiple spatial scales with the multi-camera array microscope," Optica 10, 471–480 (2023).
3. Son J, Mandracchia B, and Jia S, "Miniaturized modular-array fluorescence microscopy," Biomed. Opt. Express 11, 7221–7235 (2020). [PubMed: 33408992]
4. Fan J, Suo J, Wu J, et al. , "Video-rate imaging of biological dynamics at centimetre scale and micrometre resolution," Nat. Photonics 13, 809–816 (2019).
5. Fu Z, Jin Z, Zhang C, et al. , "The future of endoscopic navigation: a review of advanced endoscopic vision technology," IEEE Access 9, 41144–41167 (2021).
6. Göröcs Z and Ozcan A, "On-chip biomedical imaging," IEEE Rev. Biomed. Eng 6, 29–46 (2012).
7. Aharoni D, Khakh BS, Silva AJ, et al. , "All the light that we can see: a new era in miniaturized microscopy," Nat. Methods 16, 11–13 (2019). [PubMed: 30573833]
8. Kuo G, Liu FL, Grossrubatscher I, et al. , "On-chip fluorescence microscopy with a random microlens diffuser," Opt. Express 28, 8384–8399 (2020). [PubMed: 32225465]
9. Adams JK, Yan D, Wu J, et al. , "In vivo lensless microscopy via a phase mask generating diffraction patterns with high-contrast contours," Nat. Biomed. Eng 6, 617–628 (2022). [PubMed: 35256759]

10. Wu J, Chen Y, Veeraraghavan A, et al. , “Mesoscopic calcium imaging in a head-unrestrained male non-human primate using a lensless microscope,” *Nat. Commun* 15, 1271 (2024).
11. Boominathan V, Robinson JT, Waller L, et al. , “Recent advances in lensless imaging,” *Optica* 9, 1–16 (2022). [PubMed: 36338918]
12. Martínez-Corral M and Javidi B, “Fundamentals of 3D imaging and displays: a tutorial on integral imaging, light-field, and plenoptic systems,” *Adv. Opt. Photon* 10, 512–566 (2018).
13. Xue Y, Davison IG, Boas DA, et al. , “Single-shot 3D wide-field fluorescence imaging with a computational miniature mesoscope,” *Sci. Adv* 6, eabb7508 (2020).
14. Xue Y, Yang Q, Hu G, et al. , “Deep-learning-augmented computational miniature mesoscope,” *Optica* 9, 1009–1021 (2022). [PubMed: 36506462]
15. Tanida J, Kumagai T, Yamada K, et al. , “Thin observation module by bound optics (TOMBO): concept and experimental verification,” *Appl. Opt* 40, 1806–1813 (2001). [PubMed: 18357179]
16. Hu J and Yang W, “Metalens array miniaturized microscope for large-field-of-view imaging,” *Opt. Commun* 555, 130231 (2024).
17. Xu B, Li H, Gao S, et al. , “Metalens-integrated compact imaging devices for wide-field microscopy,” *Adv. Photon* 2, 066004 (2020).
18. Denis L, Thiébaud E, Soulez F, et al. , “Fast approximations of shift-variant blur,” *Int. J. Comput. Vis* 115, 253–278 (2015).
19. Debarnot V, Escande P, Mangeat T, et al. , “Learning low-dimensional models of microscopes,” *IEEE Trans. Comput. Imaging* 7, 178–190 (2020).
20. Sroubek F, Kamenicky J, and Lu YM, “Decomposition of space-variant blur in image deconvolution,” *IEEE Signal Process. Lett* 23, 346–350 (2016).
21. Yanny K, Monakhova K, Shuai RW, et al. , “Deep learning for fast spatially varying deconvolution,” *Optica* 9, 96–99 (2022).
22. Wu J, Boominathan V, Veeraraghavan A, et al. , “Real-time, deep-learning aided lensless microscope,” *Biomed. Opt. Express* 14, 4037–4051 (2023). [PubMed: 37799697]
23. He K, Zhang X, Ren S, et al. , “Deep residual learning for image recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016), pp. 770–778.
24. Szegedy C, Liu W, Jia Y, et al. , “Going deeper with convolutions,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2015), pp. 1–9.
25. Rippel O, Snoek J, and Adams RP, “Spectral representations for convolutional neural networks,” *Adv. Neural Inf. Process. Syst* 28 (2015).
26. Chen H, Huang L, Liu T, et al. , “Fourier imager network (FIN): a deep neural network for hologram reconstruction with superior external generalization,” *Light Sci. Appl* 11, 254 (2022). [PubMed: 35970839]
27. Tian F and Yang W, “Learned lensless 3D camera,” *Opt. Express* 30, 34479–34496 (2022). [PubMed: 36242459]
28. Deb D, Jiao Z, Sims R, et al. , “Fouriernet enable the design of highly non-local optical encoders for computational imaging,” *Adv. Neural Inf. Process. Syst* 35, 25224–25236 (2022)
29. Luo W, Li Y, Urtasun R, et al. , “Understanding the effective receptive field in deep convolutional neural networks,” *Adv. Neural Inf. Process. Syst* 29 (2016).
30. Yanny K, Antipa N, Liberti W, et al. , “Miniscope3D: optimized single-shot miniature 3D fluorescence microscopy,” *Light Sci. Appl* 9, 171 (2020). [PubMed: 33082940]
31. Voigt FF, Kirschenbaum D, Platonova E, et al. , “The mesospim initiative: open-source light-sheet microscopes for imaging cleared tissue,” *Nat. Methods* 16, 1105–1108 (2019). [PubMed: 31527839]
32. Todorov MI, Paetzold JC, Schoppe O, et al. , “Machine learning analysis of whole mouse brain vasculature,” *Nat. Methods* 17, 442–449 (2020). [PubMed: 32161395]
33. Hartmann J, Wong M, Gallo E, et al. , “An image-based data-driven analysis of cellular architecture in a developing tissue,” *elife* 9, e55913 (2020).
34. Srikumar T, Lewicki MC, Costanzo M, et al. , “Global analysis of sumo chain function reveals multiple roles in chromatin regulation,” *J. Cell Biol* 201, 145–163 (2013). [PubMed: 23547032]

35. Dahlin JL, Hua BK, Zucconi BE, et al. , “Reference compounds for characterizing cellular injury in high-content cellular morphology assays,” *Nat. Commun* 14, 1364 (2023).
36. Zhang Y, Li K, Li K, et al. , “Image super-resolution using very deep residual channel attention networks,” in *Proceedings of the European Conference on Computer Vision (ECCV) (2018)*, pp. 286–301.
37. Simonyan K, Vedaldi A, and Zisserman A, “Deep inside convolutional networks: visualising image classification models and saliency maps,” *arXiv*, arXiv:1312.6034 (2013).
38. Selvaraju RR, Cogswell M, Das A, et al. , “Grad-cam: visual explanations from deep networks via gradient-based localization,” in *Proceedings of the IEEE International Conference on Computer Vision (2017)*, pp. 618–626.
39. Antipa N, Kuo G, Heckel R, et al. , “Diffusercam: lensless single-exposure 3D imaging,” *Optica* 5, 1–9 (2018).
40. Hua X, Liu W, and Jia S, “High-resolution Fourier light-field microscopy for volumetric multi-color live-cell imaging,” *Optica* 8, 614–620 (2021). [PubMed: 34327282]
41. Alido J, Greene J, Xue Y, et al. , “Robust single-shot 3D fluorescence imaging in scattering media with a simulator-trained neural network,” *Opt. Express* 32, 6241–6257 (2024). [PubMed: 38439332]
42. Yang Q, Guo R, Hu G, et al. , “SV-FourierNet,” *GitHub* (2024), <https://github.com/bu-cisl/SV-FourierNet>.

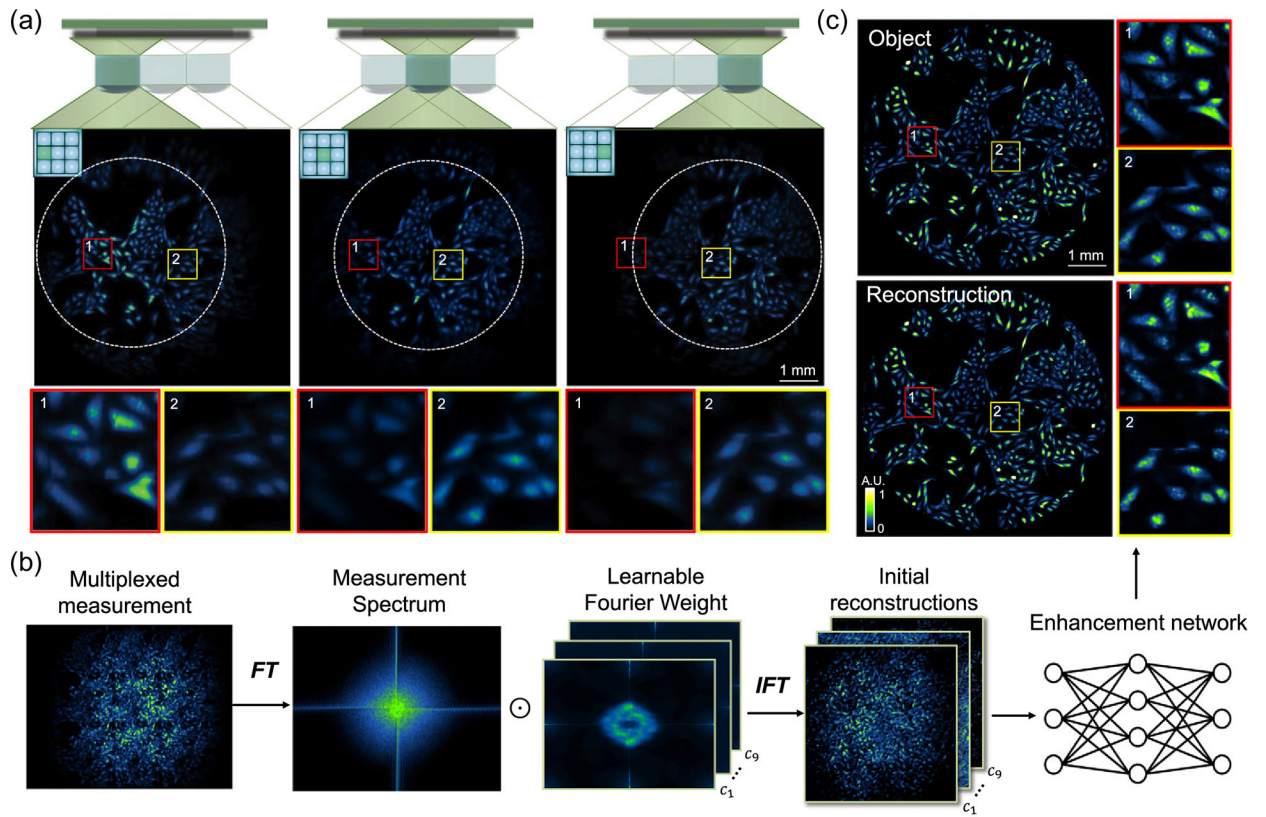


Fig. 1. System overview. (a) Illustrations of the capability of individual lenses within the MLA, each focusing on a specific sub-FOV, with noticeable aberrations and diminishing contrast at the periphery. (b) Process of SV-FourierNet, where the captured image is first transformed into the Fourier domain and processed by multiple learnable filters for initial deconvolution, followed by information fusion and enhancement by a network to yield the final output. (c) The SV-FourierNet reconstruction achieves enhanced image quality and uniform resolution across an expanded FOV.

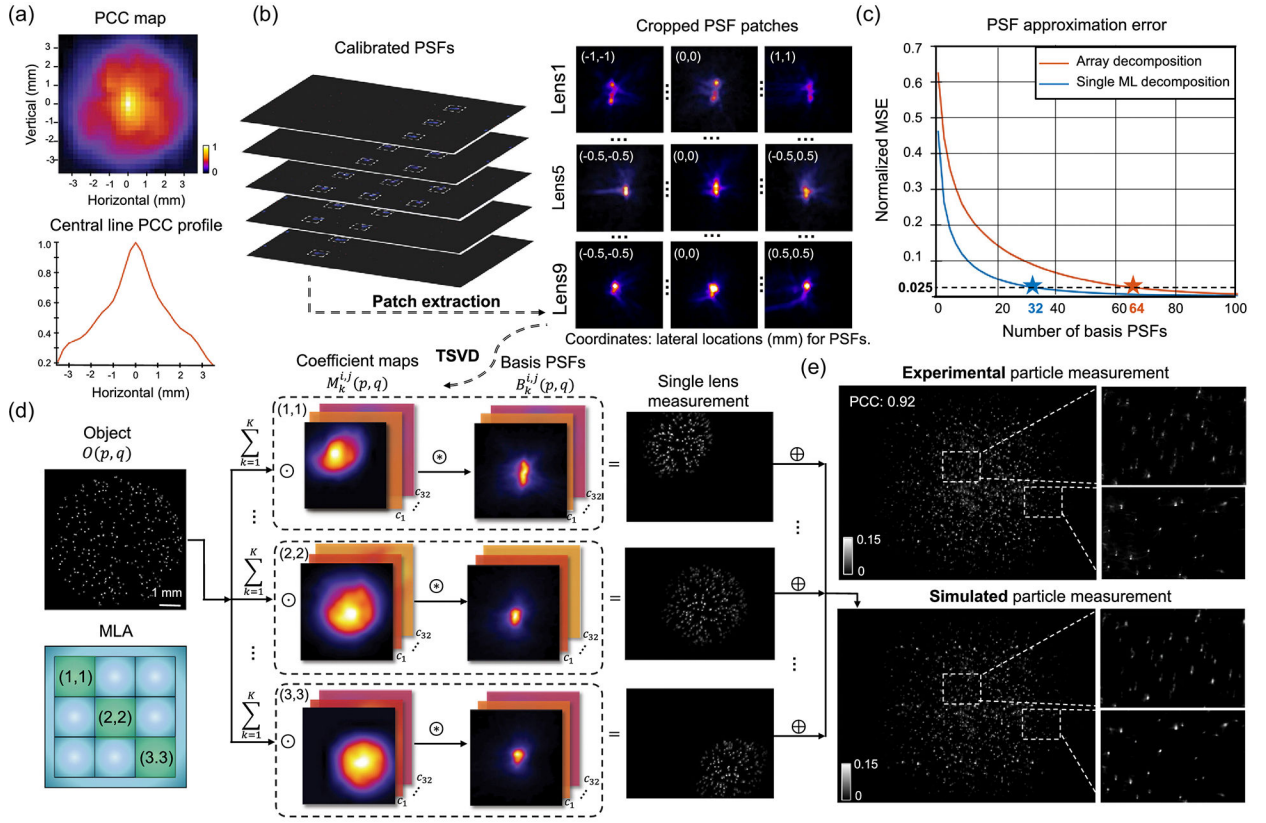


Fig. 2. Spatially varying imaging model. (a) Spatial variance is quantified by the PCC map and the central line profile, which both exhibit a marked decline as the point source shifts away from the FOV center. (b) The low-rank model is computed on calibrated PSFs, which are cropped around each ML’s foci to enhance computational efficiency. The foci captured at different point source locations under the same ML show similar aberrations. TSVD is performed to compute the basis PSFs and coefficient maps for each ML. (c) Compared to the array decomposition method, the single-ML decomposition method reduces the model’s rank by 2× while maintaining accuracy. (d) The final imaging model involves computing K weighted convolutions between the object and the basis PSFs for each ML and a superposition of all MLs to yield the multiplexed measurement. (e) The model is validated by comparing the simulated and physical measurements from the same object containing randomly distributed 10 μm fluorescent particles.

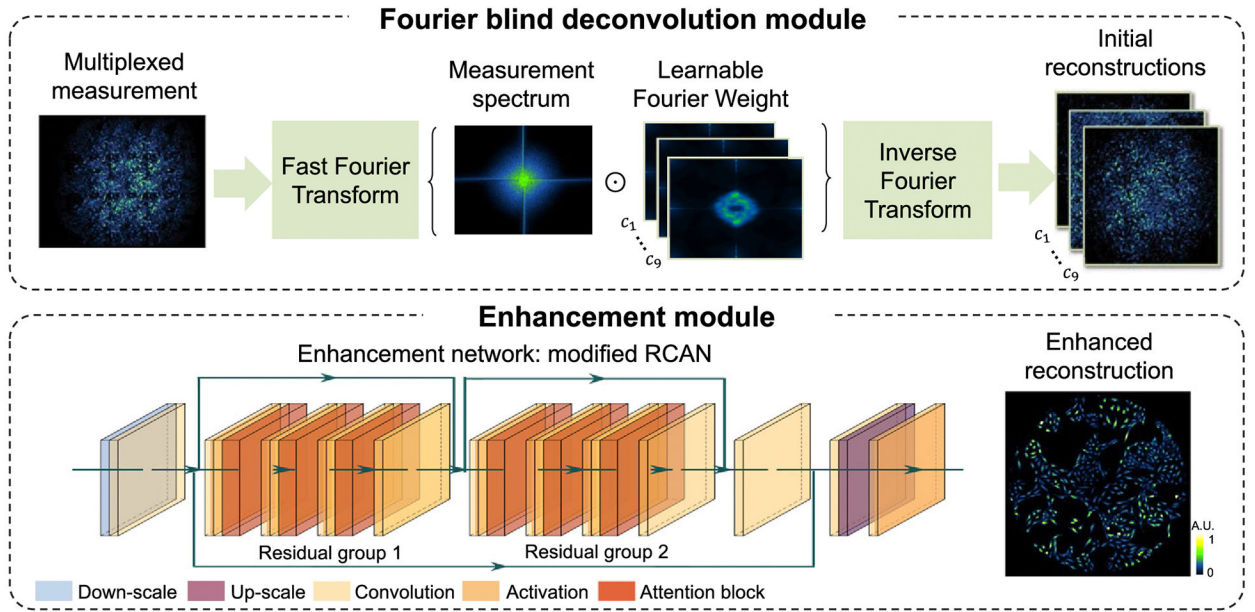


Fig. 3. SV-FourierNet structure. The full measurement is first sent to the Fourier blind deconvolution module to perform initial reconstructions, followed by the enhancement module to form the final reconstruction.

Positive ERF components V.S physical PSFs at different FOV locations

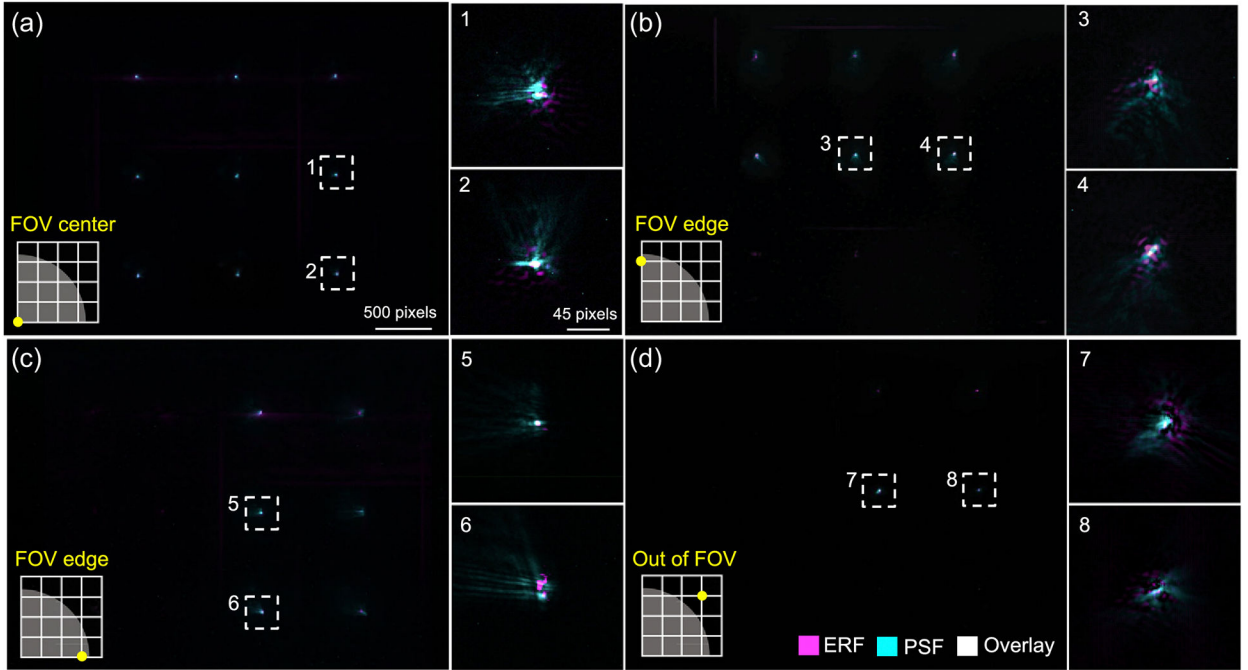


Fig. 4. SV-FourierNet’s ERFs closely match the physically measured PSFs. The figures show the overlay between SV-FourierNet’s positive ERF components (in magenta) and the physical PSFs (in cyan) at various locations: (a) the FOV center; (b), (c) near the FOV edges; and (d) beyond the training FOV. For the insets, the gray region represents a quarter of the circular FOV, marked by a grid with 1 mm intervals. The yellow marker denotes the specific position of the object point.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

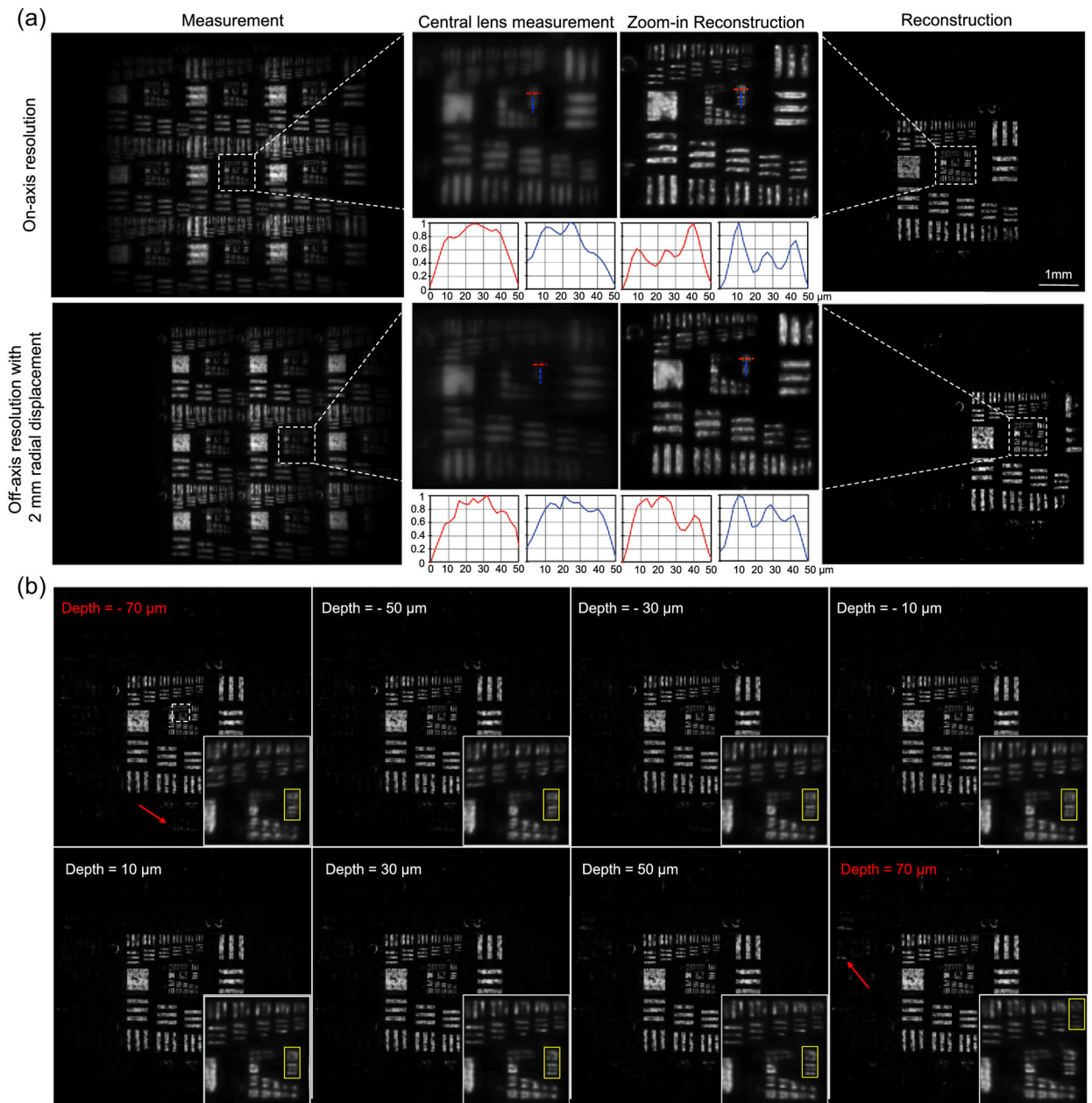


Fig. 5. Resolution characterization. (a) SV-FourierNet reconstruction of fluorescent resolution target at the FOV center and edges, compared with measurements from the central lens to highlight the enhanced resolution and extended FOV achieved by our system. Line profiles confirm the system achieves a consistent 7.8 μm resolution both vertically and horizontally, at both FOV center and edges. (b) SV-FourierNet maintains a consistent resolution within a 100 μm depth range. Beyond this range, resolution begins to degrade, and background artifacts appear, as indicated by the red arrow.

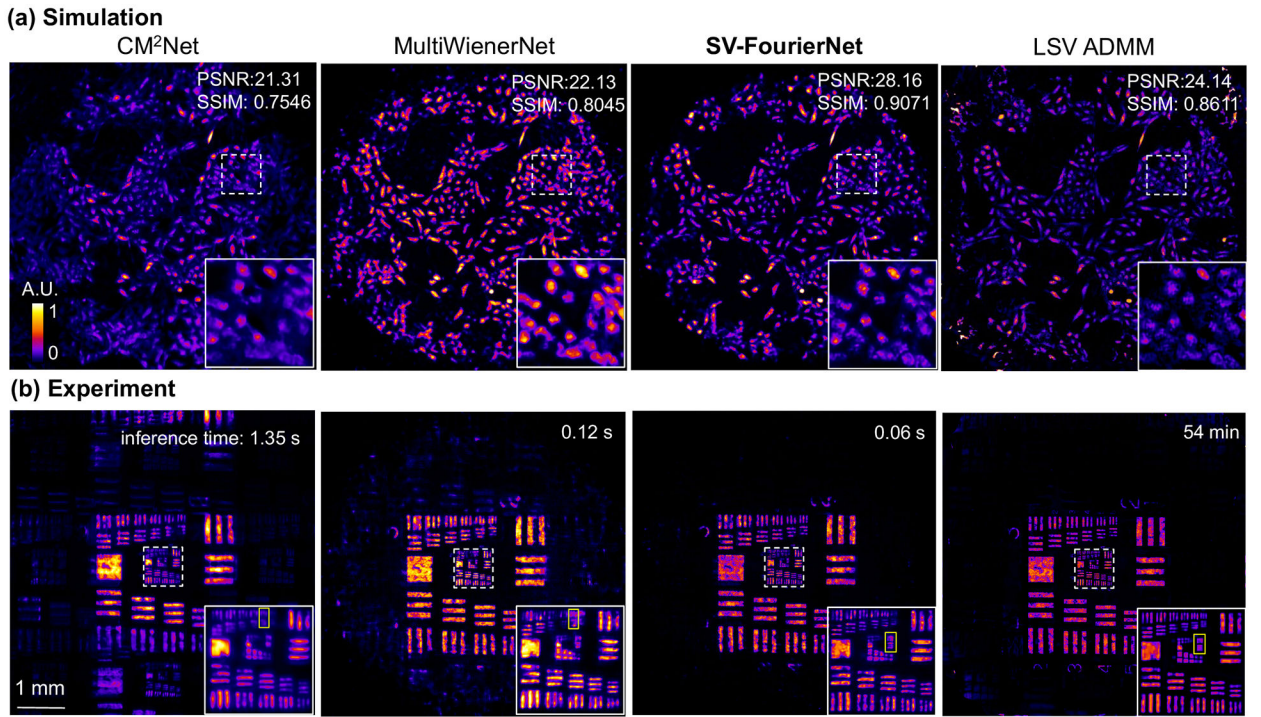


Fig. 6. Comparison of SV-FourierNet with existing state-of-the-art networks and LSV model-based algorithm (ADMM) on simulated and experimental datasets. (a) Reconstruction results from simulated cell measurement. SV-FourierNet achieves the best PSNR and SSIM. (b) Reconstruction result from experimental measurement on a resolution target. SV-FourierNet provides the best spatial resolution (highest spatial resolution achieved by each method is marked by the yellow box) and the fastest inference speed.

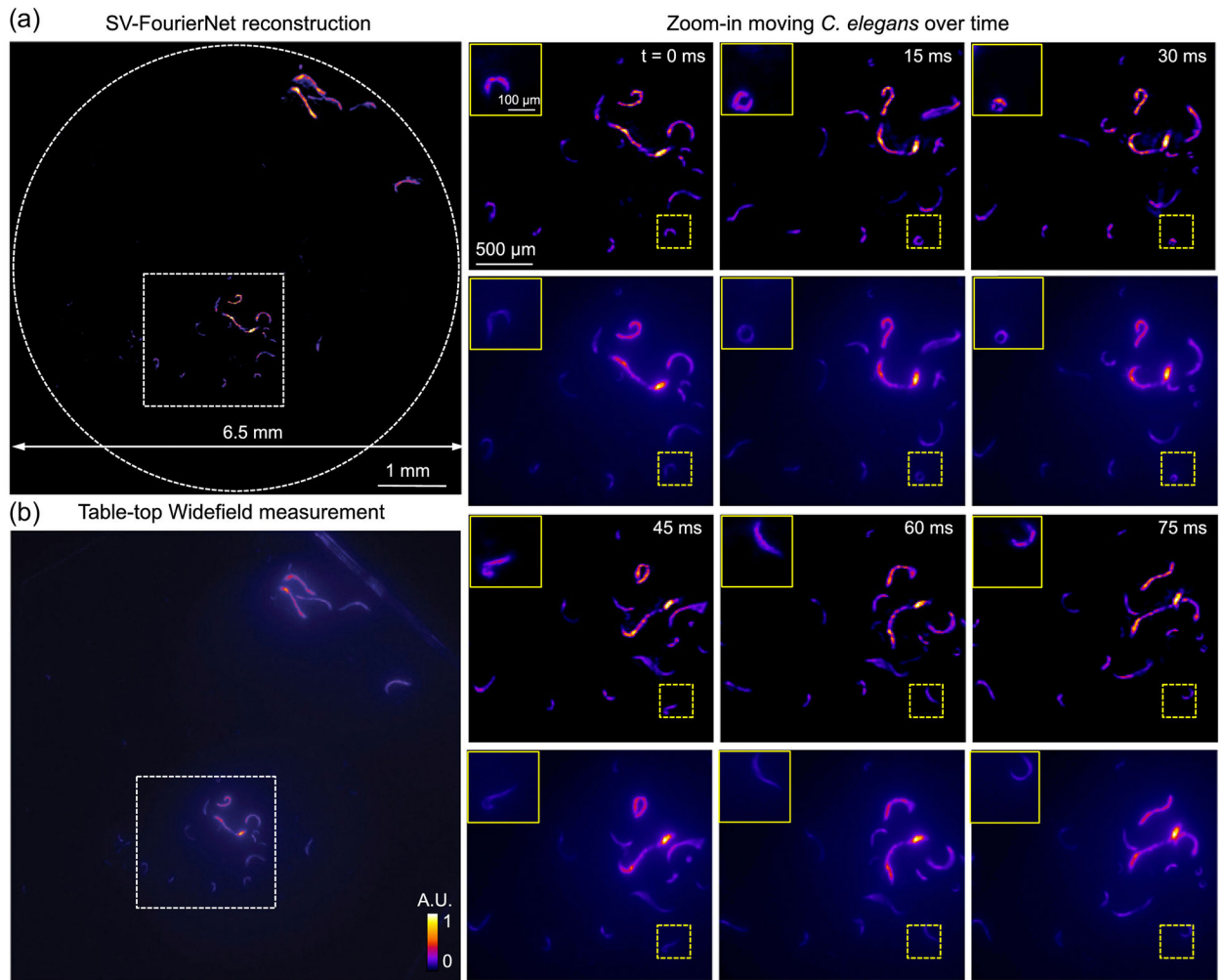


Fig. 7. Imaging of a freely moving *C. elegans* colony. (a) SV-FourierNet’s reconstruction captures the rapid movement of the *C. elegans* across a 6.5 mm FOV. (b) The reconstructions agree well with the concurrent wide-field measurements across the entire FOV and zoomed-in regions, demonstrating our system’s capacity to robustly image complex biological dynamics on a large scale.

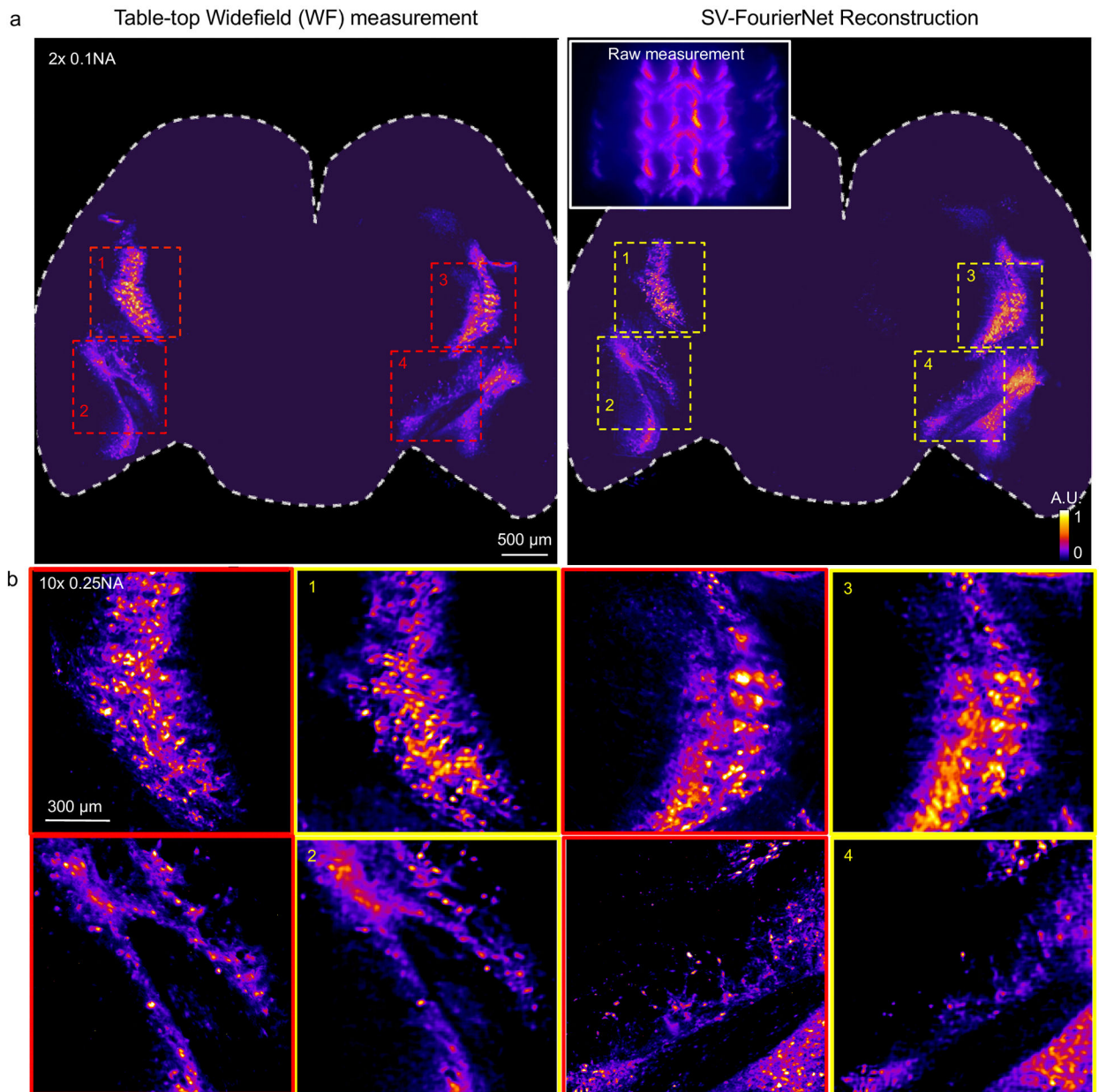


Fig. 8. Imaging of a 75 μm thick brain section. (a) SV-FourierNet reconstruction (with the raw measurement as an inset) aligns with the wide-field measurement from a $2\times 0.1\text{NA}$ objective, showcasing the network's ability for accurate reconstruction across an entire 6.5 mm brain section. (b) In the zoomed-in regions, SV-FourierNet reconstructions (with yellow image borders) agree with wide-field measurements from a $10\times 0.25\text{NA}$ objective (with red image borders), demonstrating our network's capacity to achieve reconstruction at cellular resolution.

Imaging of dynamic movements of the *C. elegans* population over a wide FOV.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript