Advanced Devices & Instrumentation
A SCIENCE PARTNER JOURNAL

RESEARCH ARTICLE

All-in-Focus Fourier Ptychographic Microscopy via 3D Implicit Neural Representation

Jingqian Wu, Chutian Wang, Shuo Zhu, Yunfei Tian, and Edmund Y. Lam*

Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam, Hong Kong SAR, China.

*Address correspondence to: elam@eee.hku.hk

Citation: Wu J, Wang C, Zhu S, Tian Y, Lam EY. All-in-Focus Fourier Ptychographic Microscopy via 3D Implicit Neural Representation. *Adv. Devices Instrum.* 2025;6:Article 0133. https://doi.org/10.34133/adi.0133

Submitted 22 June 2025
Revised 16 August 2025
Accepted 17 September 2025
Published 2 December 2025

Microscopy has long been essential to biomedical research, enabling detailed analyses of complex samples. Fourier ptychographic microscopy (FPM), a computational imaging technique, provides high-resolution, wide-field images without requiring extensive hardware modifications. However, current FPM algorithms struggle with samples exhibiting depth variations, such as tilted or 3-dimensional (3D) objects. The limited depth of field (DoF) leads to images with only focal-plane areas in sharp focus, while regions outside appear blurred. To address this limitation, we propose an all-in-focus FPM algorithm using physics-informed 3D neural representations to reconstruct sharp, wide-field images of 3D objects under limited DoF. Unlike previous methods, our approach samples the full depth range to create a 3D feature volume that incorporates spatial and depth information. A learnable 3D weight map assigns focus-based weights, emphasizing focused regions while de-emphasizing blurred ones. Additionally, we introduce an unsupervised sharpness loss to optimize the 3D heatmap, forming a new training framework for FPM that enables high-quality reconstruction without direct supervision. Experimental results on real-world microscopy data demonstrate that our approach surpasses existing methods in producing sharp, wide-field reconstructions for samples with depth variation, thus improving performance on downstream vision tasks.

Introduction

Microscopy has long played a critical role in biomedical discovery [1–4], enabling the visualization and analysis of complex biological samples. Among recent advances, Fourier ptychographic microscopy (FPM) has emerged as a powerful computational imaging technique that achieves both high resolution and wide field of view (FOV) without requiring extensive hardware modifications [5–9]. FPM captures a sequence of low-resolution intensity images under angle-varied illumination and computationally reconstructs a high-resolution image in Fourier space [10]. By enhancing the space-bandwidth product of conventional microscopy [11], FPM enables digital correction of optical aberrations [12] and facilitates wide-field, high-resolution imaging using standard microscope setups [13,14].

Recent studies have explored the use of deep-learning [15–18] frameworks to address the depth-related limitations in FPM. Implicit neural representations (INRs) have shown particular promise for continuous scene reconstruction [19–21], where a multilayer perceptron (MLP) maps spatial coordinates directly to physical values, enabling memory-efficient and high-fidelity rendering. INRs have been adapted to microscopy tasks to enhance both imaging resolution [22] and computational efficiency [23]. Furthermore, applications such as volumetric refractive index mapping in diffraction tomography [24] suggest that representing objects in 3-dimensional (3D) neural space may offer a viable path toward capturing full DoF in FPM. However, directly applying INRs to FPM remains challenging, particularly due to the absence of 3D high-resolution ground truth and the difficulty in merging volumetric information into a coherent 2D projection.

Despite its strengths, existing FPM algorithms [5,24] struggle with samples exhibiting depth variations such as thick, tilted, or inherently 3D objects. As shown in Fig. 1A, due to the limited depth of field (DoF), only regions near the focal plane remain in sharp focus, while areas outside appear blurred. When imaging objects with depth, the focal point is at the intersection (yellow star) of the optical axis and sample plane, limiting the sharp region to the area around this focus point (slashed area). Consequently, current FPM methods only yield partially sharp images, as shown in Fig. 1B, where areas outside the focal plane remain blurred.

This work presents a physics-informed all-in-focusing FPM framework that leverages 3D neural representations to reconstruct sharp, wide-field images of samples with substantial depth variation. Unlike previous approaches, our method samples the entire depth range to form a 3D feature volume incorporating both spatial and axial information. A learnable 3D weight map is introduced to assign focus-based weights, amplifying well-focused regions and suppressing blurred areas. To overcome the lack of ground-truth supervision, we design an unsupervised sharpness loss that guides the training of both the 3D representation and the weighting map. This approach offers a physically interpretable and efficient solution for depth-invariant reconstruction.

In addition to improving visual fidelity, our method has direct implications for downstream microscopy tasks such as

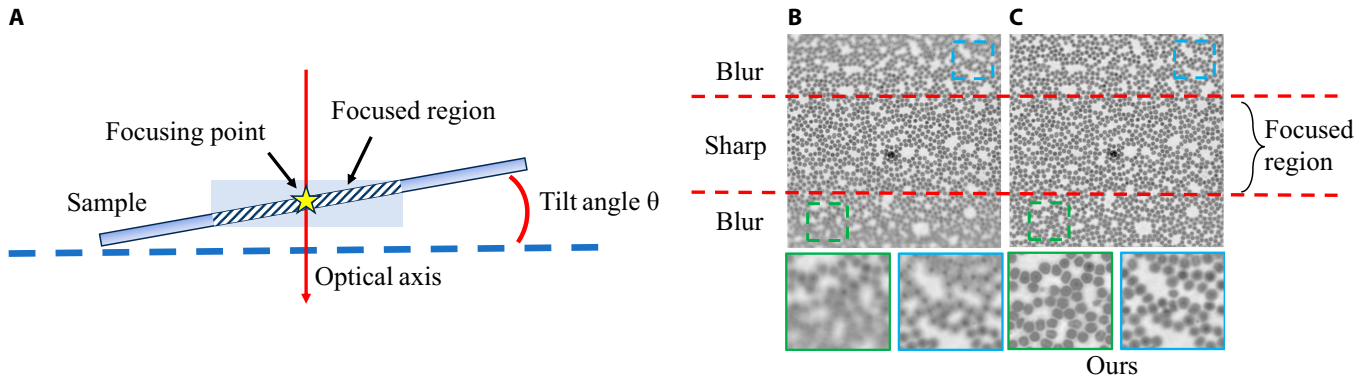

Fig. 1. (A) An illustration of imaging a 3D object under microscopy. Due to the limited depth of field, only the focused region is sharp, while areas outside the focal plane appear blurred. (B) Typical reconstruction from an existing method [24]: Only the central region is sharp, with blurred areas outside the focused region. (C) Our approach reconstructs an all-in-focus, wide-field image.

segmentation. Image segmentation is fundamental to the analysis of biological specimens, enabling cell identification, structural delineation, and morphometric quantification [25–29]. With sharp, all-in-focus reconstructions, segmentation accuracy improves substantially, particularly in high-density or multi-depth samples. As such, segmentation quality also serves as an indirect measure of FPM reconstruction performance [28].

We demonstrate our method on real-world FPM datasets, showing superior performance over state-of-the-art techniques in sharpness, structure preservation, and segmentation accuracy. By combining physical priors, 3D neural modeling, and unsupervised learning, our framework advances the capabilities of FPM in resolving complex volumetric scenes, expanding its potential in biomedical and computational imaging applications.

The remaining sections of this paper are organized as follows: Methods formally presents the problem and methodology for using neural implicit representations (INRs). Results details our experiments and analyses validating the proposed method. Finally, Conclusion concludes the paper with a discussion of limitations.

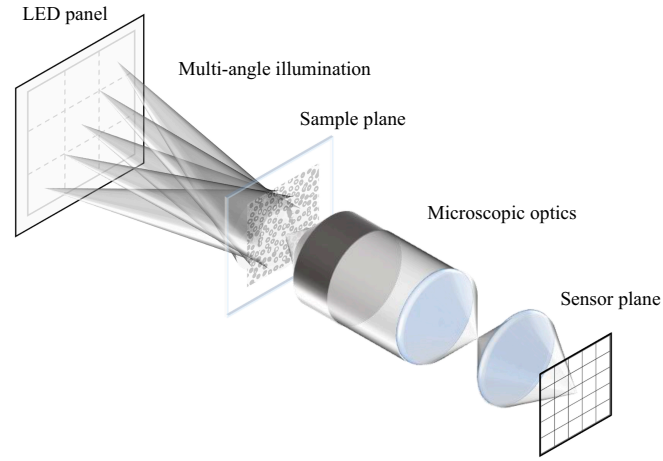

Fig. 2. Optical setup of Fourier ptychographic microscopy (FPM). A programmable LED panel provides multi-angle plane wave illumination onto the sample. The light transmitted through the sample is collected by microscopic optics and projected onto the sensor plane. Each LED position corresponds to a different incident angle, effectively shifting the sample's spatial frequency content in the Fourier domain. Sequential captures under varied illumination angles enable the reconstruction of a high-resolution, wide-field image through computational synthesis.

Methods

FPM working principle

The hardware configuration of FPM [5] is illustrated in Fig. 2. The system consists of a planar light-emitting diode (LED) array [30] as the programmable illumination source, a standard microscope objective lens, and a camera sensor [31]. During acquisition, individual LEDs in the array sequentially emit quasi-monochromatic light, forming planar waves incident on the sample from oblique angles. Each activated LED corresponds to a unique illumination direction, indicated by colored rays in the figure. The incident plane wave interacts with the 3D sample, modulating both the amplitude and phase of the transmitted wavefront [32].

This oblique illumination induces a lateral frequency shift of the sample's spectrum in the Fourier domain. As shown in the frequency inset of Fig. 2, the objective lens acts as a low-pass filter with a passband determined by its numerical aperture (NA), represented as the red circle. Oblique illumination shifts higher spatial frequency content into this limited passband, allowing the system to indirectly access spectral components beyond the conventional diffraction limit [33].

The modulated optical field passes through the objective lens and is projected onto the camera sensor, where only intensity (not phase) is recorded. By capturing a sequence of such low-resolution images under varying illumination angles, FPM effectively synthesizes a larger aperture in the Fourier domain. Through computational phase retrieval and iterative frequency stitching, a high-resolution complex image of the sample can be reconstructed.

Preliminary on FPM physics model

We first elaborate on the diffraction limit phenomenon in conventional imaging systems that occurs in the Fourier domain, which inherently restricts the optical resolution [33]. Briefly speaking, the resolution of a standard coherent imaging system is determined by the coherent transfer function (CTF) on the Fourier domain **u**. From Fourier optics, the CTF simply serves as the pupil function scaled by a factor. As depicted in Fig. 3A, for a circular aperture, the cutoff frequency of the imaging system is limited within a circle (marked in red in Fourier plane)

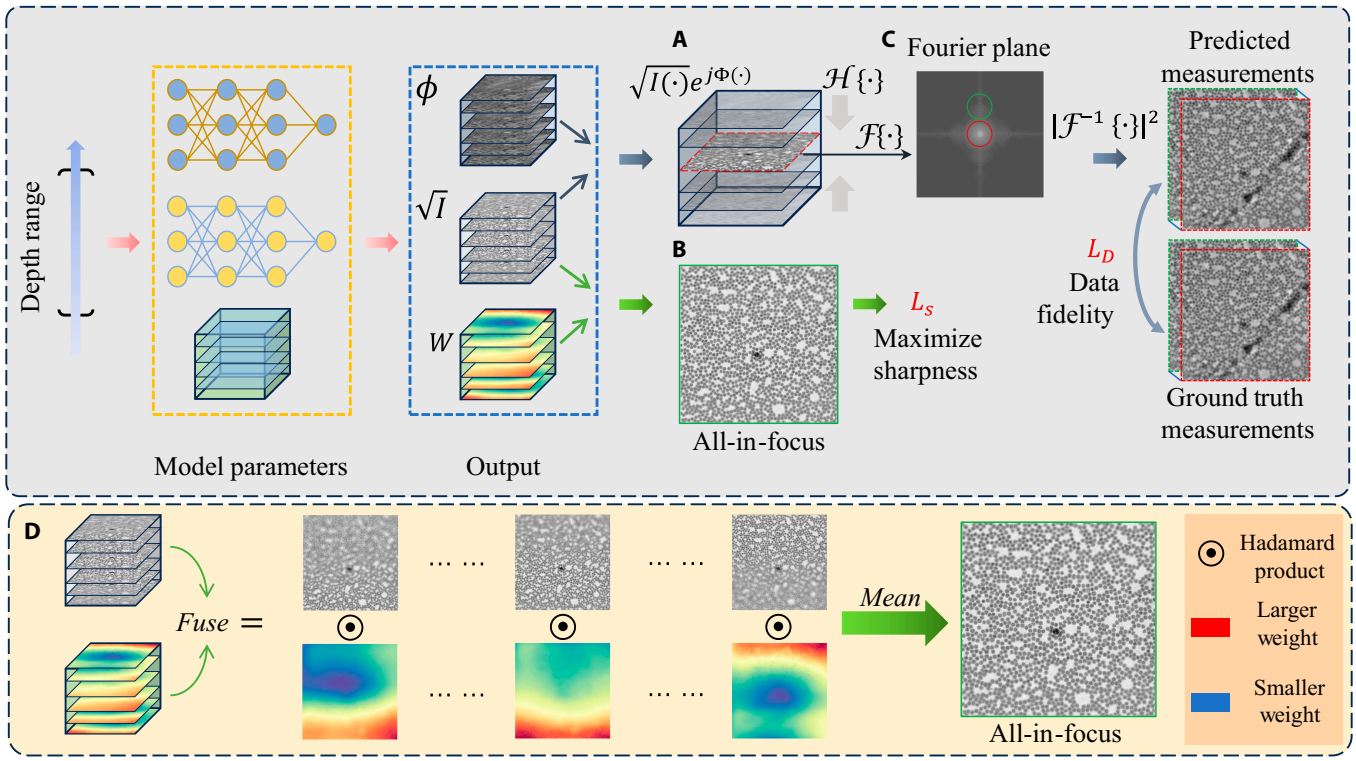

Fig. 3. Overview of the pipeline for autofocusing reconstruction under FPM. The pipeline takes in the entire focusing depth range and samples the corresponding feature space as the neural network input. The whole model parameter (yellow box) contains 2 identical MLP neural networks and a learnable weight map. The amplitude network (blue network in the yellow box) is used to predict a 3D intensity volume, and the phase network (yellow network in the yellow box) is used for a 3D phase volume. For the output, the model predicts a 3D complex phase volume ϕ and a 3D intensity amplitude volume $\sqrt{I}$, with an optimized 3D weight map. The later stages are all for supervision. The upper flow (gray arrow) denotes the process of traditional FPM data fidelity loss, where the measurements (A) are estimated by the Fourier ptychographic differentiable process (C) and will be supervised against the captured measurements using L_1 loss. The lower flow (green arrow) denotes the physically guided sharpness loss, and it is used to optimize the 3D weight map and produce the all-in-focus wide-field frame (B). (D) Our all-in-focus fusion: Each amplitude slice is weighted by the corresponding 3D map and averaged to produce the final sharp image.

$$|\mathbf{u}| < \frac{\mathrm{NA}}{\lambda}, \tag{1}$$

where $\mathbf{u} = (u, v)$ is the frequency coordinate, NA refers to the numerical aperture (half of the f-number), and λ is the illumination wavelength.

For an incoherent imaging system, the transfer function, namely, the optical transfer function (OTF), is simply the autocorrelation of the scaled pupil function. As such, although the cutoff frequency of an incoherent system can be twice of that in a coherent system, the frequency amplitude is not as flat as that of the pupil and is suffering from a descending decay that deviates from the zero frequency point.

The target in the classic FPM aims to reconstruct a super-resolution (SR) image of the sample at the focal plane that is beyond the diffraction limit. FPM involves sequentially implementing structured illumination on the sample with different illumination angles by utilizing a coded LED array.

Considering the complex optical field that describes the in-focus spatial distribution of the infocus object U_0, its expression is given by

$$U_0(\mathbf{x}) = \sqrt{I_0}(\mathbf{x})\exp\left[j\phi_0(\mathbf{x})\right], \tag{2}$$

where $\mathbf{x} = (x, y)$ serves as the spatial coordinate, I is the intensity, $\sqrt{I}$ is the amplitude, and ϕ refers to the phase of the sample. We use the subscript to simplify the z coordinate, i.e., $U_z(\mathbf{x}) = U(\mathbf{x}, z)$. Physical quantities are naturally super-resolved.

As discussed in Eq. (1), optical fields collected in the lens-based imaging systems are restricted by the scaled pupil function P, typically a circular function on the frequency domain

$$P(\mathbf{u}) = \begin{cases} 1 & \mathbf{u} < \dfrac{\mathrm{NA}}{\lambda} \\ 0 & \text{otherwise} \end{cases}. \tag{3}$$

Given the fact that the conventional camera directly captures the modular square of the complex field within the diffraction limit due to the utilized lens NA, we denote it as the low-resolution (LR) measurement

$$I_{LR}(\mathbf{x}) = \left|\mathcal{F}^{-1}\left\{\mathcal{F}\left\{U_0\right\}P(\mathbf{u})\right\}\right|^2. \tag{4}$$

When the ith monochromatic LED with a wavelength of λ illuminates the sample U, it does so at an angle $\boldsymbol{\theta}^i$ relative to the optical axis. Specifically, such an interaction on the sample field can be written as

$$U_0^i(\mathbf{x}) = \exp\left[j2\pi\frac{\sin\boldsymbol{\theta}^i}{\lambda}\mathbf{x}\right]U_0(\mathbf{x}), \tag{5}$$

where j is the imaginary unit so that the tilt illumination serves as the multiplication process between the object field and a known oblique phase term.

As illustrated in Fig. 3B, the construction of the oblique illumination manifests as a shift in the Fourier domain

$$\mathcal{F}\left\{U_0^i\right\}(\mathbf{u}) = \mathcal{F}\left\{U_0\right\}\left(\mathbf{u} - \frac{\sin\boldsymbol{\theta}^{\mathbf{i}}}{\boldsymbol{\lambda}}\right). \tag{6}$$

Therefore, while the modulated field U' carries higher frequency information along the added oblique illumination direction, the resultant intensity is enhanced by one-directional resolution, i.e.,

$$\begin{aligned} I_0^i(\mathbf{x}) &= \left|\mathcal{F}^{-1}\left\{\mathcal{F}\left\{U_0^i\right\}P(\mathbf{u})\right\}\right|^2, \\ &= \left|\mathcal{F}^{-1}\left\{\mathcal{F}\left\{U_0\right\}\left(\mathbf{u} - \frac{\sin\boldsymbol{\theta}^i}{\lambda}\right)P(\mathbf{u})\right\}\right|^2, \\ &= \left|\mathcal{F}^{-1}\left\{\mathcal{F}\left\{U_0\right\}(\mathbf{u}')P\left(\mathbf{u}' + \frac{\sin\boldsymbol{\theta}^i}{\lambda}\right)\right\}\right|^2, \end{aligned} \tag{7}$$

where $\mathbf{u}' = \mathbf{u} + \frac{\sin\boldsymbol{\theta}^i}{\lambda}$ refers to a shift of the frequency selection of the window function P, moving the truncation of the CTF to a higher frequency domain.

Suffice it to say that the oblique illumination strategy shifts back the unmeasurable higher frequency term within the diffraction limit as implied in Eq. (6). Traversing all coded illumination angle $\sin\boldsymbol{\theta}^i$ outside NA results in an extended cutoff frequency, whose maximum boundary serves as the illumination NA NA_{ill}

$$|\mathbf{u}| < \frac{\mathrm{NA} + \mathrm{NA}_{ill}}{\lambda}, \tag{8}$$

where $\mathrm{NA}_{ill} = \max\left\{\sin\boldsymbol{\theta}^i\right\}$.

As such, the captured ith LED patch can be referred to as super-resolved supervision I_{SR}^i within the specific region (red circle) in the Fourier plane.

Capturing extended frequency in one plane already requires numerous measurements, which extremely restricts the acquisition efficiency of multiple depth planes. Fortunately, in the (quasi)-coherent imaging system, the propagation between different planes obeys the basic scalar diffraction theory in Fourier optics. This allows the multiplexing of the focal plane's measurements as the supervision of different depths, i.e., yielding a reconstruction of a volumetric field rather than just a planar one. To be specific, for the complex field U_z at arbitrary depth z, we physically back-propagate the field to the focal plane $z = 0$ and implement the supervision. For simplicity, the physical back-propagation process is denoted as

$$U_0(\mathbf{x}) = \mathcal{H}\left\{U_z(\mathbf{x}), -z\right\}, \tag{9}$$

where $\mathcal{H}$ serves as the free-space propagation operator to transport field U_{z_0} from z_0 to z_n. It can be written as

$$\mathcal{H}\left\{U_{z_0}; z_n - z_0\right\} = \mathcal{F}^{-1}\{\mathcal{F}\left\{U_{z_0}\right\}(\mathbf{u})\times \exp \tag{10}$$

3D feature volume with depth information

INRs have recently emerged as a powerful paradigm for modeling continuous signals by learning a mapping from coordinates (e.g., 3D position) to signal values using neural networks. When applied to volumetric data, the network learns a continuous representation in a 3D neural space, enabling high-fidelity reconstruction without discretized grids or voxel storage.

Our goal is to reconstruct an all-in-focus wide-field frame by using the focus depth range $z_n \in \left[z_0, z_{N_z}\right]$, instead of a single focus depth z. Therefore, we sample the object under the microscope in 3D space. To represent a 3D volumetric sample, we initialize 2 identical 3D feature volumes $\mathcal{V} \in \mathbb{R}^{H\times W\times N_z} = \left\{v_{i,j}\right\}$. Specifically, the spatial coordinates (x, y) are discretized into a regular grid covering the FOV, while the entire set of depth coordinates $\left\{z_n\right\}_{n=0}^{N_z}$ is normalized to $[-1, 1]$. We then simultaneously sample features across all depth planes, producing the complete 3D feature volume $V \in \mathbb{R}^{H,W,N_z} = \left\{v_{i,j}\right\}$, where $i, j \in \mathbb{R}^{H,W}$, and each feature v on the spatial map holds a learnable feature vector $v \in \mathbb{R}^{N_z}$.

By default, the number of depth layers N_z used during training is calculated as follows:

$$\Delta z = 0.8 \times \mathrm{DOF}, \quad N_z = \left\lceil \frac{z_{\max} - z_{\min}}{\Delta z} \right\rceil \tag{11}$$

where DOF denotes the system's depth of field, and $z_{\min}, z_{\max}$ define the axial range of the sample.

However, this configuration is not fixed. Our approach allows for freely specifying the number of depth layers, N_z, depending on the axial complexity of the sample. Our method is designed to generalize well to deeper layers. By using the continuous depth range representation, the model can adapt to denser or sparser sampling along the depth axis.

Two real-valued MLPs are contained in the model (yellow box in Fig. 3), and each takes a 3D feature volume V independently to generate phase and amplitude components. Specifically, they are the phase network (top blue network in the yellow box), which is used to generate the predicted 3D phase volume $\left\{\phi_{z_n}\right\}_{H\times W\times N_z}$, and the amplitude network (middle yellow in the yellow box), which is used to generate the expected 3D square-root intensity volume $\left\{\sqrt{I}_{z_n}\right\}_{H\times W\times N_z}$. Using V as input, each neural network is trained to map the vectors $\in \mathbb{R}^{N_z}$ in V to values in the field of different focus distance $z_n, n \in N_z$. Each MLP begins with N_z channels in its first layer, followed by 2 nonlinear layers with ReLU activations and a final linear layer to output the 3D super-resolved intensity and phase volume. Bilinear interpolation along the spatial dimension addresses gaps in pixel count, yielding a compact neural representation that requires only a few thousand parameters, speeding up reconstruction.

Physics-informed focus weight map

To perceive the focus region, illustrated in Fig. 1A, we employ a weight map W (bottom yellow box in Fig. 3). The key motivation behind this physically informed weight map is that it explicitly tells, at different levels of focus depth z_n, which region is the focus region and which region is the defocused region. Just as demonstrated in Fig. 4, when we explicitly sample a tilted object (dark blue in Fig. 4A) in 3D space, we need the weight map (3D volume in Fig. 4B) to perceive where the corresponding focus regions at different focus depths, and the focus regions (dark blue rectangles in Fig. 4B) should be assigned larger weights, where the defocused regions should be assigned smaller

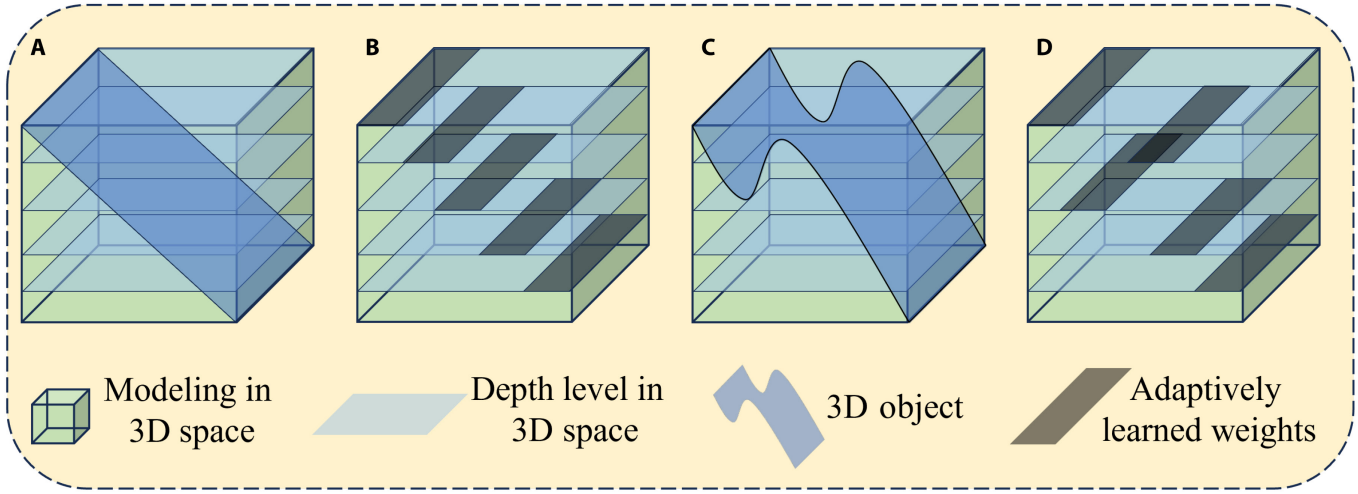

Fig. 4. Illustration of tilt or 3D sample under microscope. (A) Dark blue as a tilted sample in 3D space, with a light green plane on each height at a different focusing distance. (B) Black areas at different planes as focusing areas at different focusing distances, and where the 3D weight map should be maximized. (C and D) Different illustration for random 3D object with random depth pattern compared to (A) and (B).

weights. Please note that here we demonstrate the weight map in a discrete way, but in the actual learning process, it should be a continuous 3D tensor.

To generate one single spatial all-in-focus wide-field frame, we rely on the trained weight map W (bottom yellow box in Fig. 3). The network predicts a 3D amplitude volume $\sqrt{I} \in \mathbb{R}^{H\times W\times N_z}$, where each slice corresponds to the intensity response at a different focal depth z_n. To synthesize the final 2D output image Y, we perform a weighted fusion across the depth dimension by taking a per-pixel weighted mean along z:

$$Y = \frac{1}{N_z}\sum_{n=0}^{N_z} \mathrm{Softmax}_z\{W\}(z_n) \circ \sqrt{I}(z_n), \tag{12}$$

where $\mathrm{Softmax}_z\{\cdot\}$ refers to a 1D Softmax operation on the z axis for $H \times W$ channels. This process adaptively selects the most in-focus regions from each depth slice based on the learned focus-aware weights. As illustrated in Fig. 3D, this fusion mechanism allows us to retain only the sharpest structures from different depth layers, thereby producing a clean, all-in-focus 2D projection that eliminates depth-dependent blurring. Such a strategy mimics the human manual focusing process but in a fully data-driven and differentiable way, enabling end-to-end optimization during training.

Physically informed FPM supervision

Data fidelity loss

To get the supervision from the measured super-resolved image stack, we use 2 neural networks to represent the SR amplitude and the SR phase at the imaging depth z, i.e.,

$$\begin{aligned} f_1(z;\boldsymbol{\theta}_1) &= \sqrt{I}_z(\mathbf{x}), \\ f_2(z;\boldsymbol{\theta}_2) &= \phi_z(\mathbf{x}), \end{aligned} \tag{13}$$

where $z = 0$ is set as the origin on the focal plane, and $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_1$ refer to the parameters of the 2 neural networks f_1 and f_2, respectively.

The object complex field can therefore be expressed as

$$U_z(\mathbf{x};\boldsymbol{\theta}) = f_1(z;\boldsymbol{\theta}_1)\exp\left[jf_2(z;\boldsymbol{\theta}_2)\right]. \tag{14}$$

As aforementioned in Eq. (9), for any depth layer in the stack V, a physical back-propagation process is implemented to multiplex the supervision on the focal plane. Subsequently, the neural network representation of the back-propagated field in Eq. (14) is integrated into Eqs. (6) and (7), yielding the all-depth data fidelity loss term in FPM

$$L_D = \sum_{n=0}^{N_z}\sum_{i=0}^{N_{\mathrm{ill}}} \left\| \left|\mathcal{H}\{U^i_{z_n}(\mathbf{x};\boldsymbol{\theta}); -z_n\}\right|^2 - \hat{I}^i_0(\mathbf{x}) \right\|_1, \tag{15}$$

where N_{ill} refers to the number of the illumination angle and I^i is the ith recorded obliquely illuminated sample intensity. Herein, we use smoothL1 loss in PyTorch to guarantee data fidelity.

Physically guided sharpness loss

As shown in Fig. 3, the all-in-focus wide-field frame Y (sub-image with the green box in Fig. 3) is generated using the square root of the amplitude I and the weight map W. However, as explained in Introduction, one of the major challenges is to design an effective supervision strategy that encourages sharpness across the entire FOV.

Regularization techniques are widely used in computational imaging to guide neural network optimization with meaningful priors and to improve robustness under ill-posed or unsupervised settings [34–36]. Such priors can be structural (e.g., sparsity, smoothness, and edge-preserving constraints) and are especially valuable when ground truth data are unavailable. Inspired by this, we adopt a physically interpretable sharpness regularization by computing the L1-norm of the gradient on the rendered all-in-focus frame Y (left red box in Fig. 3), following the principle of sparsity of gradient (SOG) [37]. This encourages sharper reconstructions by emphasizing high-frequency spatial variations. The full sharpness loss is defined in Eq. (12) and jointly optimizes the learnable weight map W to favor in-focus regions, i.e.,

$$L_S = -\kappa_S{}^*|\nabla Y(\mathbf{x})|_1, \tag{16}$$

where κ_S is the weight of the sharpness regularizer and $\nabla Y(\mathbf{x}) = \left(\frac{\partial Y}{\partial x}, \frac{\partial Y}{\partial y}\right)^T$ refers to the gradient. In our implementation,

the coefficient κ_S is empirically selected to balance the contribution of the sharpness regularization relative to the data fidelity loss. As a form of regularization, this coefficient serves to control how strongly the model emphasizes sharpness during optimization. A small κ_S might make the model ignore sharpness cues, while an overly large value could over-constrain the solution and suppress useful features. We started from a moderate value and conducted sensitivity analyses to ensure stable training and robust reconstructions across datasets. This kind of tuning is consistent with prior works in computational imaging and inverse problems, where regularization parameters are often adjusted based on the optimization landscape and task-specific requirements. Importantly, we found that the performance was not overly sensitive to the precise value of κ_S within a reasonable range. We use $\kappa_S = 0.5$ for all experiments.

This simple yet effective supervision allows our approach to accurately optimize the physical-guided weight map described in Physics-informed focus weight map. The optimized result can be found in Fig. 5. Figure 5A and C shows the selected planes with specific focus depth z_1 and z_2. Due to different focus depths, for a tilted 3D object, the focused region would be different. For example, Fig. 5A can be focused in the middle, where the top and bottom remain blurred. In this case, the optimized heatmap with corresponding heatmap (Fig. 5B) would maximize (shown in red) the weight value in the middle region and minimize (shown in blue) the weight in the remaining regions. The same applies to Fig. 5C, where the focused region is at the bottom, so the optimized heatmap will provide large weights (red) to the bottom. As a result, rendering the all-in-focus frame Y using the 3D volume $\left\{\sqrt{I}\right\}$ and 3D heatmap W produces sharp and focused regions. By adopting the physics-informed focus weight map and the physically guided sharpness loss, our algorithm does not require any real 3D volume object as a label for learning. This enables learning directly from raw FPM measurements while exploiting the inherent physical constraints of light propagation.

As the proposed physically guided sharpness loss provides effective supervision, in scenarios, however, where the sample exhibits naturally low contrast or minimal textural variation (e.g., uniformly stained tissue or homogeneous cytoplasmic regions), the gradient magnitude differences between focused and defocused regions may be diminished. This could, in principle, reduce the effectiveness of our sharpness-based weighting scheme. To mitigate this limitation, we designed our framework to jointly learn a 3D focus-aware weighting map across the depth volume, introduced in Physics-informed focus weight map. Rather than relying on local gradient magnitude alone, the optimization process encourages global consistency across layers, guided by the physical forward model and overall reconstruction fidelity. This makes our approach more robust than purely local sharpness heuristics.

Results

Dataset and experimental setup

To demonstrate the effectiveness of our approach, we use the dataset collected by [24]. We categorize all data samples into 2 types. (a) Tilted samples to represent 3D objects with depth under the microscope: Here, we demonstrate that our approach is able to reconstruct an all-in-focused wide-field frame, whereas existing approaches are only able to image partially focused frames. (b) Plane samples with no depth involved: Here, we demonstrate that existing approaches require a human selection process of picking the right focus depth to produce a focused wide-field frame and filtering out all other wrong-defocused frames. Our approach is agnostic to the accuracy of

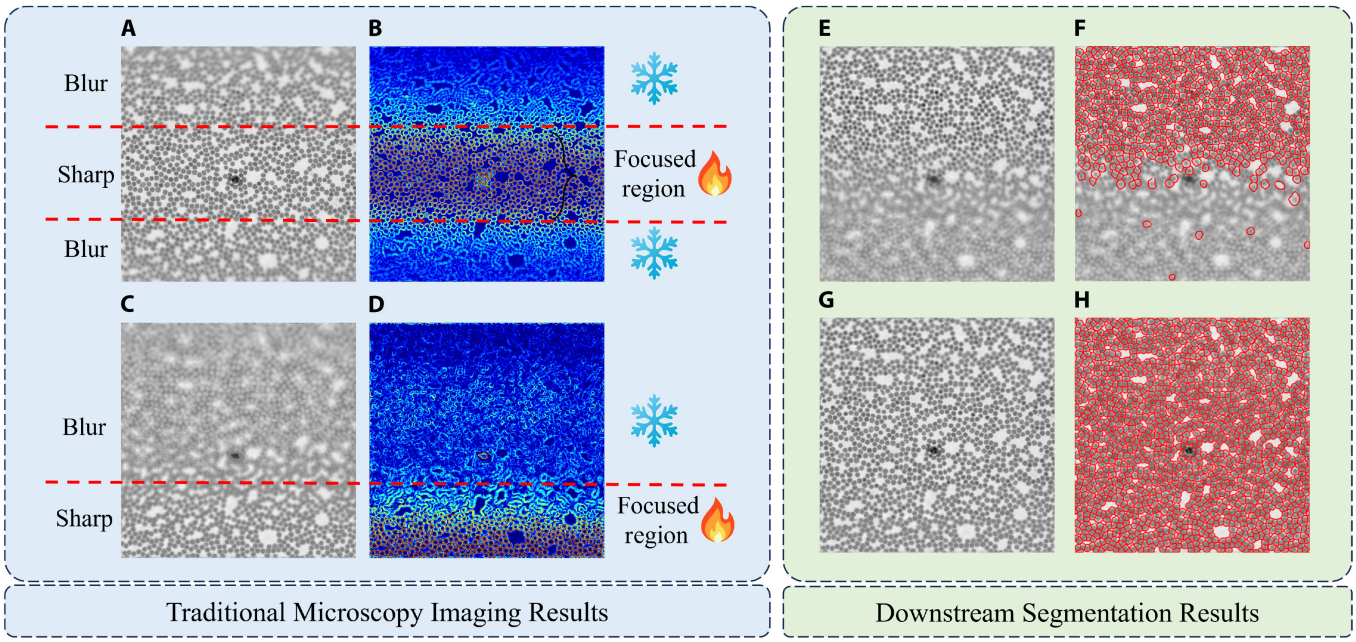

Fig. 5. Blue box: Illustration of optimizing the weight map for a tilted 3D sample under a microscope. The optimized weight map (B and D) produces a high probability (red) in the sharp region in the corresponding reconstructed frame with different focus depths (A and C) and a low probability (blue) in the blurred region. This helps us to produce a sharp all-in-focus wide-field frame, as described in Eq. (12). Green box: Comparison of segmentation results using the Cellpose [27] on different outputs. (E) Reconstruction from [24] where a partial region is focused and the rest remains blurred and defocused. (F) Cellpose segmentation on [24], where blurriness leads to poor segmentation quality. (G) Reconstruction for 3D objects from ours that is all-in-focused and sharp across all regions. (H) Cellpose segmentation on our output, showing clear and precise boundaries.

focus depth and produces the focused sharp wide-field frame given a range of input focus depths. For all experiments, we compare with (a) reconstruction results from the classic physics FPM approach [5] by picking the most focused frame with focus depth z in all focus depth range z_0 to z_{N_z}; (b) the reconstruction results from the most recent neural-based FPM [24], with random focus depth $z = z_n$; (c) reconstruction results from picking the most focused frame with focus depth z in all focus depth range z_0 to z_{N_z} from [24]; and (d) ours.

3D tilted samples

We used a tilted human blood smear slide to validate our proposed approach on 3D samples. The slide was tilted at an angle of 4° relative to the optical axis of the microscope. Illumination was provided by a 32 × 32 LED array (Adafruit, 4-mm pitch) and a 16-element LED ring, with an illumination NA matched to the objective lens NA (Olympus PLN 10×/0.25 NA). A total of 68 LEDs were used for sequential illuminations and imaging at a central wavelength of 522 μm. The LED panel was

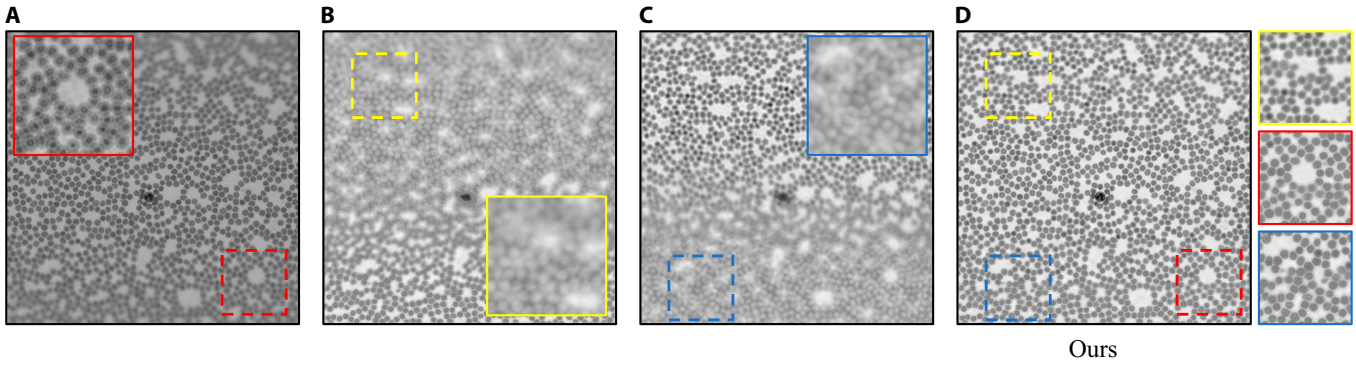

Fig. 6. Visualization comparison of our approach and existing FPM approaches. (A) Reconstruction result from [5]. (B) Reconstruction result from [24] with randomly selected depth value. (C) Reconstruction result from [24] with human-selected best depth value. (D) Our reconstruction result without human selection. Our approach is able to reconstruct an all-in-focus wide-field frame with clear and sharp results in different regions.

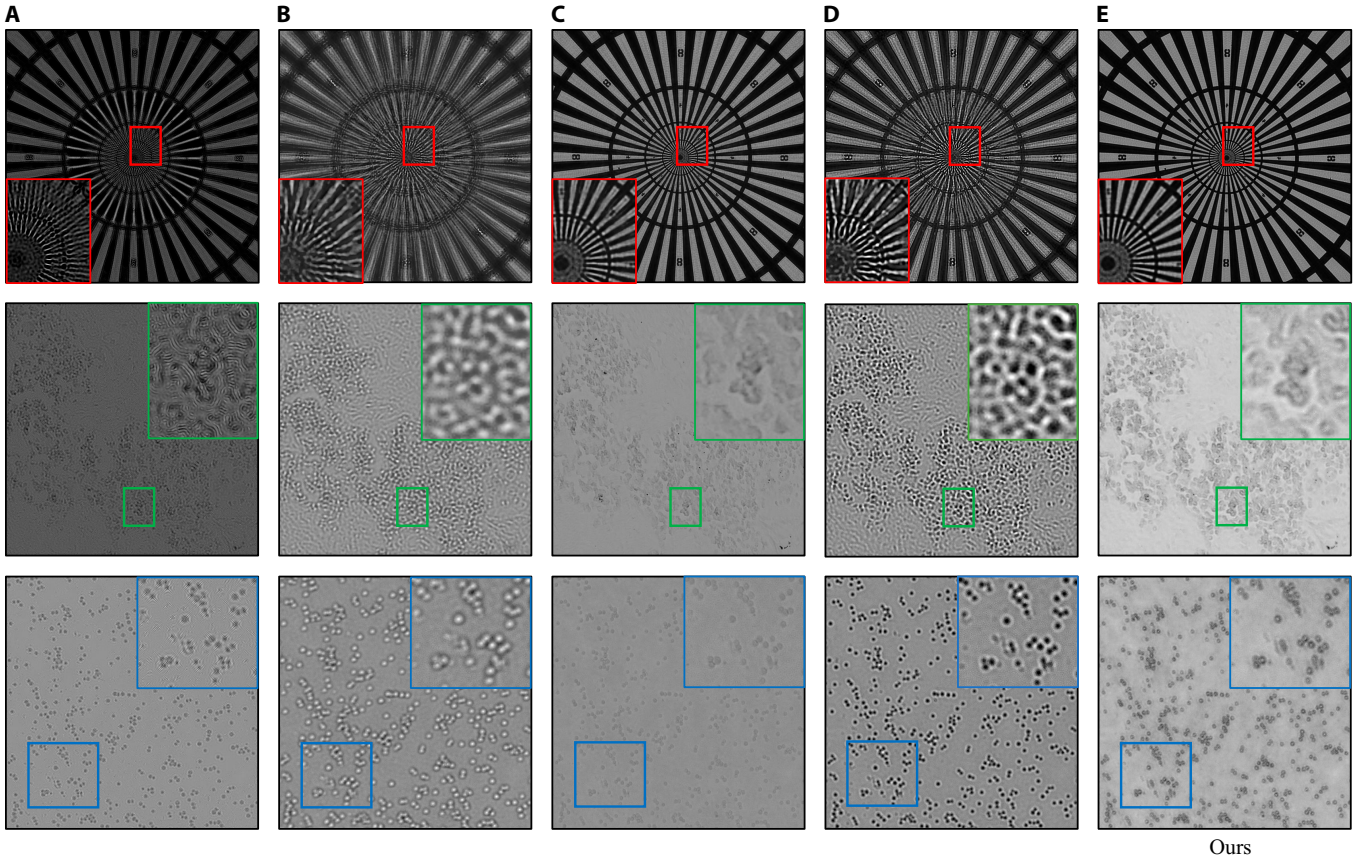

Fig. 7. Visualization comparison of our approach and existing FPM approaches. (A) Reconstruction result from [5]. (B) Reconstruction result from [24] with randomly selected depth value. (C) Reconstruction result from [24] with human-selected best depth value. (D) Reconstruction result from [38]. (E) Our reconstruction result without human selection. Our approach is able to reconstruct a sharp, wide-field frame without selecting the correct focus depth.

positioned 74 mm from the sample. A monochrome camera (Allied Vision Prosilica GT6400) with a pixel pitch of 3.45 μm was used to capture images. All components were assembled and mounted on an Olympus IX51 inverted microscope body.

To compare visually against existing FPM approaches, we demonstrate the visualization comparison in Fig. 6. Figure 6A shows the physics FPM approach with the best focus depth picked by humans: Although most middle regions seem to be sharp and clear, the bottom and top regions still remain a blur. Panels (B) and (C) of Fig. 6 are both from [24]. Figure 6D shows the reconstruction results from a classical autofocusing approach [38]. The 4 reconstruction results from existing approaches remain blurred in different regions. However, ours (Fig. 6D) is able to reconstruct an all-in-focus wide-field frame with sharp results across different regions.

3D plane samples

We also conduct experiments on plane objects under the microscope. Current FPM algorithms require a focus depth z in focus depth range z_0 to z_{N_z}. Therefore, it is critical to find the correct and suitable focus depth z to produce the focus reconstruction result. Otherwise, the reconstruction result will be defocused and blurry. However, our approach samples the entire focus depth range into 3D space and reconstructs a sharp wide-field frame using our proposed physically guided weight map, where the defocused frames are assigned less weight. Figure 7 demonstrates the visual comparison between our approach (Fig. 7E) and existing approaches described previously.

Downstream tasks: Segmentation

An important aspect of microscopy imaging is that it provides further analysis and downstream applications. The quality of downstream applications also depends on the reconstruction quality of the algorithm. Therefore, to further validate the effectiveness of our proposed method, we assess its utility in computer vision downstream tasks, specifically image segmentation, as it enables the precise delineation of structures within complex biological samples, allowing researchers to identify and analyze distinct cellular or tissue components. Using the Cellpose segmentation framework [27], we compare segmentation results on our reconstruction result with those generated by [24].

In Fig. 5, panels (E) and (G) present reconstructions of 3D objects from [24] and ours, respectively. Our reconstruction (Fig. 5G) demonstrates uniform focus and sharpness across all regions, whereas [24] (Fig. 5E) displays a partial region in focus, with surrounding areas blurred and defocused. We applied the Cellpose segmentation algorithm on both reconstructions of the blood smear cell, hoping we could segment out each individual cell on a blood smear imaging result under a microscope. Cellpose's output on [24] reconstruction, shown in Fig. 5F, reveals a massive wrong-segmented area due to blurriness and defocused areas. In contrast, Cellpose's output on our result, shown in Fig. 5H, provides clear, well-defined boundaries, indicative of effective segmentation across all regions. By providing an all-in-focus, high-resolution representation, our approach supports downstream tasks in vision, such as segmentation, where clarity across the entire image is essential for optimal performance. These results underscore the potential of our method to enhance interpretability and accuracy in microscopy-based vision applications.

Conclusion

We proposed a physics-informed all-in-focus FPM algorithm that reconstructs sharp, wide-field images of 3D samples, overcoming depth-of-field limitations faced by existing methods. Our approach addresses limitations in traditional FPM approaches by capturing depth information in a learnable 3D feature volume and introducing a differentiable heatmap to dynamically emphasize focused regions across depth layers. We leverage unsupervised sharpness loss, enabling it to learn effectively without requiring direct supervision. Results demonstrate that our approach outperforms existing FPM techniques, achieving enhanced image sharpness and resolution across complex sample configurations. This framework offers a practical solution for microscopy applications where depth variation is a significant challenge, particularly in biomedical imaging.

The limitation of our work is that the explored data samples in this paper are limited. Specifically, too few 3D samples are obtained and used. Therefore, for future work, we would extend our work by introducing a new FPM hardware imaging system. Another challenge is to obtain a corresponding wide-field all-in-focused ground truth frame for such a 3D sample. Further efforts would also be investigated in this direction. We aim to collect and launch a novel dataset for rendering all-in-focused 3D samples with ground truth under the microscope to contribute to this research community.

Acknowledgments

Funding: This work is supported by Research Grants Council of Hong Kong (GRF 17201620) and Theme-based Research Scheme (T45-701/22-R).
Author contributions: J.W. conceived the idea, designed the method, performed the experiments, and wrote the manuscript. C.W. helped algorithm implementation and data processing. S.Z. assisted in experimental setup and validation. Y.T. contributed to data analysis. E.Y.L. supervised the project, provided guidance, and revised the manuscript.
Competing interests: The authors declare that they have no competing interests.

Data Availability

The data/code supporting the findings of this study are available from the corresponding author upon reasonable request. Please contact E.Y.L. at elam@eee.hku.hk or J.W. at jingqianwu99@gmail.com for access to the relevant data.

References

1. Albers L, Hagemann M, Schnabel R. Laser scanning microscopy for tomographic imaging of roughness and point absorbers in optical surfaces. *Adv Dev Instrum*. 2024;5:0064.
2. Wu J, Zhu S, Wang C, Shi B, Lam EY. SweepEvGS: Event-based 3D Gaussian splatting for macro and micro radiance field rendering from a single sweep. *IEEE Trans Circuits Syst Video Technol*. 2025.
3. Zhang Y, Lam EY. Robust holographic imaging for real-world applications with joint optimization. *Opt Express*. 2025;33(3):5932–5944.
4. Ye Z, Huang Y, Zhang J, Chen Y, Ye H, Ji C, Jin L, Gan Y, Sun Y, Tao W, et al. Universal and high-fidelity resolution

extending for fluorescence microscopy using a single-training physics-informed sparse neural network. *Intel Comput*. 2024;3:0082.
5. Zheng G, Horstmeyer R, Yang C. Wide-field, high-resolution Fourier ptychographic microscopy. *Nat Photonics*. 2013;7(9):739–745.
6. Ou X, Horstmeyer R, Yang C, Zheng G. Quantitative phase imaging via Fourier ptychographic microscopy. *Opt Lett*. 2013;38(3):4845–4848.
7. Nguyen T, Xue Y, Li Y, Tian L, Nehmetallah G. Deep learning approach for Fourier ptychography microscopy. *Opt Express*. 2018;26(20):26470–26484.
8. Zhang Y, Jiang W, Tian L, Waller L, Dai Q. Self-learning based Fourier ptychographic microscopy. *Opt Express*. 2015;23:18471–18486.
9. Ye X, Sun J, Jiang W, Yu R, Chen C, Xiao X, Qian X, Huang C, Hu Y, Zhu S, et al. Ultracompact multimode meta-microscope based on both spatial and guided-wave illumination. *Adv Dev Instrum*. 2023;4:0023.
10. Guo Y, Liang Y, Li Y, Tian B, Fan X, He Y, Liu M, Peng L, Tang N, Tan T, et al. Optical microcavities empowered biochemical sensing: Status and prospects. *Adv Dev Instrum*. 2024;5:0041.
11. Bian L, Suo J, Dai Q, Chen F. Fourier ptychography for high space-bandwidth product microscopy. *Adv Opt Technol*. 2017;6:449–457.
12. Zhou A, Wang W, Chen N, Lam EY, Lee B, Situ G. Fast and robust misalignment correction of Fourier ptychographic microscopy for full field of view reconstruction. *Opt Express*. 2018;26(18):23661–23674.
13. Zheng G, Shen C, Jiang S, Song P, Yang C. Concept, implementations and applications of Fourier ptychography. *Nat Rev Phys*. 2021;3(3):207–223.
14. Song L, Wang K, Zhu S, Ren Z, Lam EY. Ptychographic image reconstruction with a phase-assisted alternating direction method of multipliers. *Opt Lett*. 2025;50(9):2816–2819.
15. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436–444.
16. Goodfellow I, Bengio Y, Courville A. *Deep learning*. Cambridge (UK): MIT Press; 2016.
17. Tong Y, Pivnenko M, Chu D. Implementation of 10-bit phase modulation for phase-only LCOS devices using deep learning. *Adv Dev Instrum*. 2020;2020:9515747.
18. Li C, Rai MR, Cai Y, Ghashghaei HT, Greenbaum A. Intelligent beam optimization for light-sheet fluorescence microscopy through deep learning. *Intell Comput*. 2024;3:0095.
19. Sitzmann V, Martel J, Bergman A, Lindell D, Wetzstein G. Implicit neural representations with periodic activation functions. *Adv Neural Inf Proces Syst*. 2020;33:7462–7473.
20. Zhou H, Shen C, Liang M, Yang C. Analysis of postreconstruction digital refocusing in Fourier ptychographic microscopy. *Opt Eng*. 2022;61(07):73102.
21. Zheng C, Zhang S, Zhou G, Hu Y, Hao Q. Robust Fourier ptychographic microscopy via a physics-based defocusing strategy for calibrating angle-varied LED illumination. *Biomed Opt Express*. 2022;13(3):1581–1594.
22. Wang H, Zhu J, Li Y, Yang Q, Tian L. NeuPh: Scalable and generalizable neural phase retrieval with local conditional neural fields. *Adv Photon Nexus*. 2024;3(03):56005.
23. Yang R, Xiao T, Cheng Y, Li A, Qu J, Liang R, Bao S, Wang X, Wang J, Suo J, et al. Sharing massive biomedical data at magnitudes lower bandwidth using implicit neural function. *Proc Natl Acad Sci*. 2024;121(28): Article e2320870121.
24. Zhou H, Feng BY, Guo H, Lin S, Liang M, Metzler CA, Yang C. Fourier ptychographic microscopy image stack reconstruction using implicit neural representations. *Optica*. 2023;10(12):1679–1687.
25. Xing F, Xie Y, Su H, Liu F, Yang L. Deep learning in microscopy image analysis: A survey. *IEEE Trans Neural Netw Learn Syst*. 2017;29:4550–4568.
26. Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, Xiao T, Whitehead S, Berg AC, Lo W-Y, et al. Segment anything. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. Vancouver (Canada): IEEE; 2023. p. 4015–4026.
27. Stringer C, Wang T, Michaelos M, Pachitariu M. Cellpose: A generalist algorithm for cellular segmentation. *Nat Methods*. 2021;18(1):100–106.
28. Liu Z, Jin L, Chen J, et al. A survey on applications of deep learning in microscopy image analysis. *Comput Biol Med*. 2021;134:104523.
29. Wu J, Zhu S, Wang C, Lam EY. Ev-GS: Event-based gaussian splatting for efficient and accurate radiance field rendering. In: *2024 IEEE 34th International Workshop on Machine Learning for Signal Processing (MLSP)*. London (UK): IEEE; 2024. p. 1–6.
30. Liu C, Li B, Qiu M. Advancements in the improvement of optical outcoupling efficiency for perovskite LEDs. *Adv Dev Instrum*. 2024;5:0045.
31. Chen N, Wu Y, Tan C, Cao L, Wang J, Lam EY. Uncertainty-aware Fourier ptychography. *Light Sci Appl*. 2025;14(1):236.
32. Wang K, Song L, Wang C, et al. On the use of deep learning for phase recovery. *Light Sci Appl*. 2024;13:4.
33. Goodman JW. *Introduction to Fourier optics*. 4th ed. New York (NY): Roberts and Company Publishers; 2005.
34. Wu J, Yang F, Cao L. Resolution enhancement of long-range imaging with sparse apertures. *Opt Lasers Eng*. 2022;155:107068.
35. Lee KC, Chae H, Xu S, Lee K, Horstmeyer R, Lee SA, Hong B-W. Anisotropic regularization for sparsely sampled and noise-robust Fourier ptychography. *Opt Express*. 2024;32(14):25343–25361.
36. Zhang S, Pan A, Sun H, Tan Y, Cao L. High-fidelity computational microscopy via feature-domain phase retrieval. *Adv Sci*. 2025;21(21):2413975.
37. Zhang Y, Wang H, Wu Y, Tamamitsu M, Ozcan A. Edge sparsity criterion for robust holographic autofocusing. *Opt Lett*. 2017;42(19):3824–3827.
38. Geusebroek JM, Cornelissen F, Smeulders AW, Geerts H. Robust autofocusing in microscopy. *Cytometry*. 2000;39(1):1–9.