

ARTICLE

Open Access

Multispectral extended depth-of-field fluorescence microscopy with co-designed meta-optics and neural reconstruction

Ipek Anil Atalay Appak^{1,2}, Haobijam Johnson Singh¹ , Sanna Korpela³, Teemu O. Ihalainen³, Erdem Sahin⁴, Christine Guillemot² and Humeysra Caglayan^{1,5} 

Abstract

High numerical aperture fluorescence microscopy provides subcellular resolution, but its depth of field is extremely limited, so thick specimens quickly fall out of focus and typically require axial scanning. Multispectral imaging further compounds this problem because chromatic aberrations shift the best focus plane and distort registration across emission channels, especially in thick samples. In this work, we present MANTIS (Multispectral All-Depth meta-optic Imaging System), a co-designed computational microscopy imaging system that achieves extended depth-of-field from a single acquisition without axial scanning. MANTIS combines a learned meta-optic with a physics-guided neural network trained end-to-end to reconstruct sharp multispectral images from depth, and wavelength-dependent blurred sensor measurements. We experimentally demonstrate a 50 μm extended depth of field at NA 1.1, corresponding to an 82-fold increase over a conventional wide-field microscope. In addition, simulations show that MANTIS can target different depth-of-field ranges, with the expected trade-off that larger depth-of-field ranges come at a cost in reconstruction fidelity. We validate the approach on biologically relevant fluorescence specimens, including 50 μm thick three-dimensional cultured MDCK II spheroids, and show that, compared with conventional wide-field fluorescence microscopy, reconstructions maintain contrast and lateral detail across depth, with reduced defocus blur and consistent performance across spectral channels.

Introduction

Fluorescence microscopy is a cornerstone imaging technique in biological and biomedical research. It works by labeling specific molecules in a sample with fluorescent dyes that emit light upon excitation at a specific wavelength. This enables precise, high-contrast visualization of objects such as individual proteins, organelles, and entire cells. To achieve detailed images at the sub-cellular levels, fluorescence microscopes often use high numerical aperture (NA) objectives. A high NA lens can collect more light and resolve finer details. However, this increase in resolution comes with a fundamental limitation: a

significantly reduced depth-of-field (DoF), the axial range over which objects appear sharp in an image. As NA increases, DoF shrinks rapidly, often to just a few microns or less¹. This means that only a thin slice of a 3D biological sample is sharply in focus at any given time, while structures closer to or further away from the in-focus region appear blurred. This shallow DoF is particularly problematic for thick or heterogeneous samples, such as tissue sections, organoids, or even single cells with complex internal structures. In live-cell imaging, the issue is compounded by the fact that the sample can be dynamic and fast-moving, requiring rapid imaging across multiple planes.

Traditionally, microscopes mitigate shallow DoF by axial scanning, acquiring a z-stack by stepping the specimen or the objective along the optical axis. The slices are then fused to produce an all-in-focus image. Confocal

Correspondence: Humeysra Caglayan (h.caglayan@tue.nl)

¹Physics Unit, Faculty of Engineering and Natural Sciences, Tampere University, Tampere 33720, Finland

²INRIA Rennes, Bretagne Atlantique Research Centre, Rennes 35042, France

Full list of author information is available at the end of the article

© The Author(s) 2026



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.



microscopy provides optical sectioning by rejecting out-of-focus light with a pinhole. Still, it relies on horizontal and axial scanning to build a volume or to synthesize extended-depth-of-field (EDoF) views. Imaging by scanning increases acquisition time and is prone to motion artifacts in live specimens. Repeated illumination further increases photobleaching and phototoxicity, which limits throughput and temporal resolution. A range of advanced optical techniques has been developed to reduce reliance on mechanical z-scanning, most notably light-sheet microscopy, remote focusing, and spatial or spectral multiplexing. In light-sheet microscopy^{2,3}, illumination and detection are decoupled so that only a thin slice of the sample is excited, minimizing out-of-focus background. Remote focusing^{4–6} shifts the focal plane by tuning optical power rather than physically moving components. Spatial and spectral multiplexing^{7–9} simultaneously acquire multiple planes or channels. Despite their strengths, these methods often require elaborate hardware with additional lenses, beam splitters, or custom illumination paths, increasing both cost and system complexity. Many implementations still rely on sequential acquisition across depths or wavelengths, preventing full-volume capture at the camera's native frame rate. Two-photon microscopy offers deeper penetration but relies on high-peak-intensity pulses, which increase photobleaching and phototoxicity^{10,11}, limiting its suitability for live-cell experiments.

Multispectral fluorescence imaging adds an additional challenge on top of the EDoF constraint. The microscope point spread function (PSF) depends on NA, wavelength, and defocus, so any residual chromatic aberration causes the best-focus plane and PSF width to vary slightly between channels. In thick specimens this leads to depth-dependent misregistration: at a given plane some channels are sharp while others are blurred, producing color fringes and cross-channel offsets that complicate co-localization and quantitative analysis across the axial range. A practical system should therefore maintain nearly uniform resolution and accurate cross-channel registration over wavelength and depth from a single exposure per field of view (FoV), without resorting to axial stacking. Although modern high-NA objectives strongly suppress chromatic aberrations, residual wavelength dependence of the PSF and dispersive elements elsewhere in the optical path can still introduce non-negligible chromatic shifts in thick, multi-color samples.

Computational imaging offers a powerful route to multispectral EDoF by jointly designing the optics and the reconstruction algorithm. In wavefront-coding approaches, the PSF is deliberately engineered so that sensor measurements remain informative across a large depth range and can be computationally restored^{12–18}. Although effective, most implementations to date rely on

diffractive optical elements (DOEs), whose inherent dispersion complicates chromatic correction. This makes it challenging to manage wavelength-dependent aberrations at the physical layer, particularly when multiple spectral channels must be captured simultaneously with high precision.

Meta-optics offers a path beyond these limitations. By providing subwavelength control over phase, dispersion, and polarization within an ultrathin platform, meta-optics enable optical functions far exceeding those of conventional DOEs^{19,20}. Leveraging this flexibility, several EDoF metalens architectures combine engineered phase profiles with computational reconstruction to achieve broadband, depth-invariant imaging. Examples include rotationally symmetric EDoF metalenses paired with deconvolution, fast varifocal meta-optics, and inverse-designed lenses that generate uniform focal spots or lens-like PSFs across extended depths^{21–25}. These studies demonstrate that meta-optics with post-processing can handle RGB or broadband scenes under defocus. However, the reported designs use moderate NA and are typically evaluated on macroscopic, planar targets, so the EDoF corresponds to relatively mild blur. Consequently, these approaches do not yet enable high NA volumetric multispectral imaging of thick biological specimens in a single shot at high spatial resolution. Parallel advances in achromatic metalens arrays have demonstrated snapshot full-color light-field and integral imaging, offering broadband, multi-view capture within a single exposure^{26–28}. However, unlike single-aperture EDoF designs, array-based approaches divide the finite pixel and photon budget across multiple views, inherently reducing per-view resolution and signal-to-noise ratio (SNR).

These developments are particularly relevant in fluorescence microscopy, where spatial resolution, spectral fidelity, and DoF are tightly coupled. Thick, multi-label biological samples demand accurate color registration across depth, while high NA amplify defocus blur differently across different wavelength channels. As a result, existing EDoF meta-optic systems, though promising, still fall short for practical volumetric fluorescence imaging, often requiring axial stacks, multi-plane acquisition, or reduced NA to maintain image quality. The convergence of meta-optics and computational imaging now opens the door to overcoming these limitations. By jointly tailoring the physical wavefront transformation and the reconstruction algorithm, it becomes possible to encode depth and wavelength-dependent information in a single measurement and decode it computationally without sacrificing spatial resolution.

Here we realize this concept in MANTIS, the Multispectral All Depth meta-optic Imaging System, a co-designed computational microscopy imaging system that



addresses both EDoF and chromatic aberration and enables high-resolution multispectral three-dimensional fluorescence imaging from a single shot. MANTIS combines a learned meta-optic with a deblurring convolutional neural network (D-CNN) that is trained together within a physics-based differentiable forward model with a stable multi-term loss, allowing optimization of the full imaging pipeline without axial scanning, color corrected relay optics, or per channel calibration. This co-optimization yields a compact and high-speed system that performs volumetric multispectral reconstruction from a single exposure. In experiments, three-dimensional fluorescence images collected with MANTIS closely track the physics-based simulations, supporting the robustness of the learned design. Together, these results establish MANTIS as a multispectral EDoF imaging approach that delivers high resolution, cross-channel registered volumetric reconstructions in a single acquisition without mechanical scanning or additional corrective optics.

Results

To realize MANTIS, we developed an integrated optical and computational system that jointly addresses chromatic aberration and defocus blur. The overall architecture and learning pipeline are illustrated in Fig. 1.

The optical setup (Fig. 1a) is based on a modified wide-field fluorescence microscope. To access the native Fourier plane within the objective, a secondary 4f relay system is introduced to reimage it externally. An optimized meta-optic, implemented as a TiO₂ nanopillar array, is positioned at the reimaged Fourier plane to modulate the wavefront before image formation. The captured sensor images are subsequently processed by a trained D-CNN to reconstruct high-resolution sample images across multiple depths and wavelengths.

Figure 1b illustrates the differentiable end-to-end learning framework used to optimize the entire imaging system. During training, high-resolution image patches, annotated with their corresponding depth and spectral channels, are propagated through a physics-based forward model that emulates image formation at the sensor plane. The simulated sensor outputs are then processed by the D-CNN, which is trained to reconstruct the original object. The meta-optic phase profile and network parameters are jointly optimized via backpropagation, enabling co-adaptation of optical encoding and computational decoding toward optimal reconstruction fidelity. After training convergence, the optimized meta-optic is fabricated and integrated into the fluorescence microscope to enable single-shot multispectral EDoF imaging. Further implementation details are provided in the following sections.

Simulations

Our framework, MANTIS, is developed to address the inherent trade-off between the DoF and spatial resolution in multispectral fluorescence imaging. This relationship is described by

$$\text{DOF} \propto \frac{\lambda}{\text{NA}^2} \propto \frac{\text{Resolution}^2}{\lambda} \quad (1)$$

indicating that achieving a large DOF becomes more challenging with high NA objectives and shorter wavelengths. We use a wide-field fluorescence microscope with a 1.1 NA objective, making the problem particularly demanding.

In simulations, all optical parameters are fixed and identical across experiments (see Section 1.1 in Supplementary). The only configuration-dependent change is the Fourier-plane sampling grid Δs , which is fixed by two criteria: (i) the space-bandwidth product (SBP) required by the passed object-space bandwidth, (ii) and a phase-sampling bound set by the worst defocus at the shortest wavelength. For incoherent imaging, the required space-bandwidth product is given by:

$$\begin{aligned} \text{SBP}(\lambda) &\approx \sum_{\lambda} \left[(2L_x f_{\text{pass}}(\lambda)) (2L_y f_{\text{pass}}(\lambda)) \right], \\ f_{\text{pass}}(\lambda) &= \min\{f_c(\lambda), f_{\text{Nyq,obj}}\}, L_{x,y} = (N_{x,y}p)/M \end{aligned} \quad (2)$$

where $f_c = 2\text{NA}/(\lambda)$ is the incoherent cutoff spatial frequency, $f_{\text{Nyq,obj}} = 1/(2p/M)$ is the camera Nyquist frequency mapped to the object space, $L_{x,y}$ is the object-space FoV, $N_{x,y}$ the sensor pixel counts, p is the sensor pixel size, and M is the magnification. For our system, $\text{SBP}(\lambda)$ is 4.2M, 3.3M, 2.6M, 2.05M at {433, 521, 601, 681} nm, totaling $\text{SBP}_{\text{total}} = 12.3\text{M}$.

Furthermore, to prevent PSF aliasing, we impose an upper bound on the Fourier sampling grid. The critical Fourier sampling grid, Δs_{crit} , is bounded by the maximum phase slope at the pupil edge, which is determined by the maximum defocus at the shortest wavelength¹⁴:

$$\Delta s_{\text{crit}} = \frac{\pi(D/2)}{8\psi_{\text{max}}(\lambda_{\text{min}})}, \psi_z = \frac{\pi}{\lambda} \left(\frac{1}{z} - \frac{1}{z_0} \right) (D/2)^2 \quad (3)$$

where ψ_z is the defocus coefficient, D is the entrance pupil diameter, z_0 is the nominal focus and $\psi_{\text{max}} = \max(|\psi_z|, |\psi_{z^+}|)$ over the target DOF. We then oversample this critical grid value so that the resulting Fourier sampling pitch supports the required total SBP across all wavelengths and reduces phase error and numerical



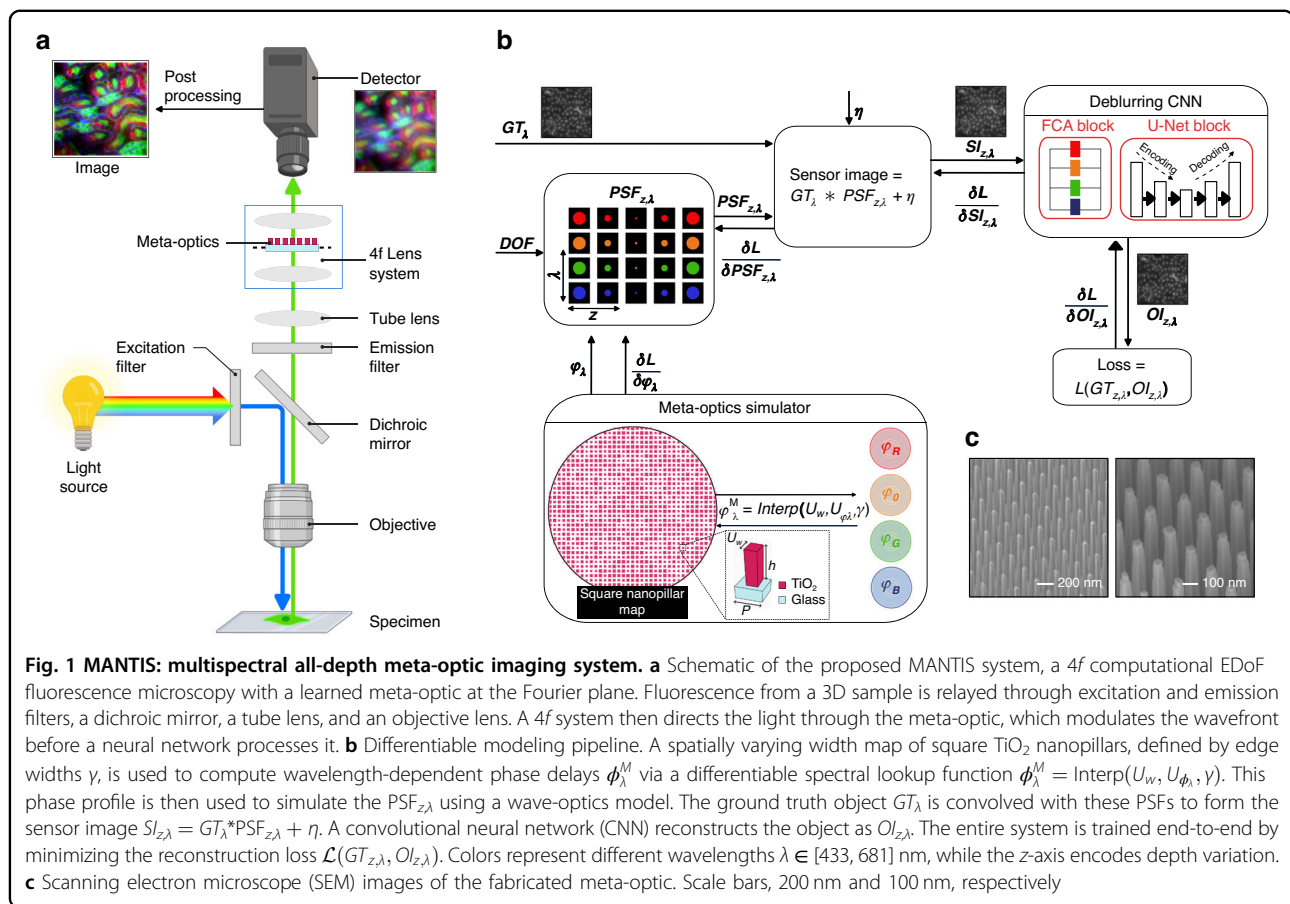


Table 1 Quantitative comparison of trained configurations targeting different EDoF ranges

Target DoF	ψ_{max}	Δs (μm)	PSNR (dB)	SSIM
25 μm	110.20	2.25	24.99	0.78
50 μm	221.32	1.125	23.53	0.70
75 μm	333.38	0.675	22.51	0.68

PSNR and SSIM values are averaged over depth and spectral channels

aliasing:

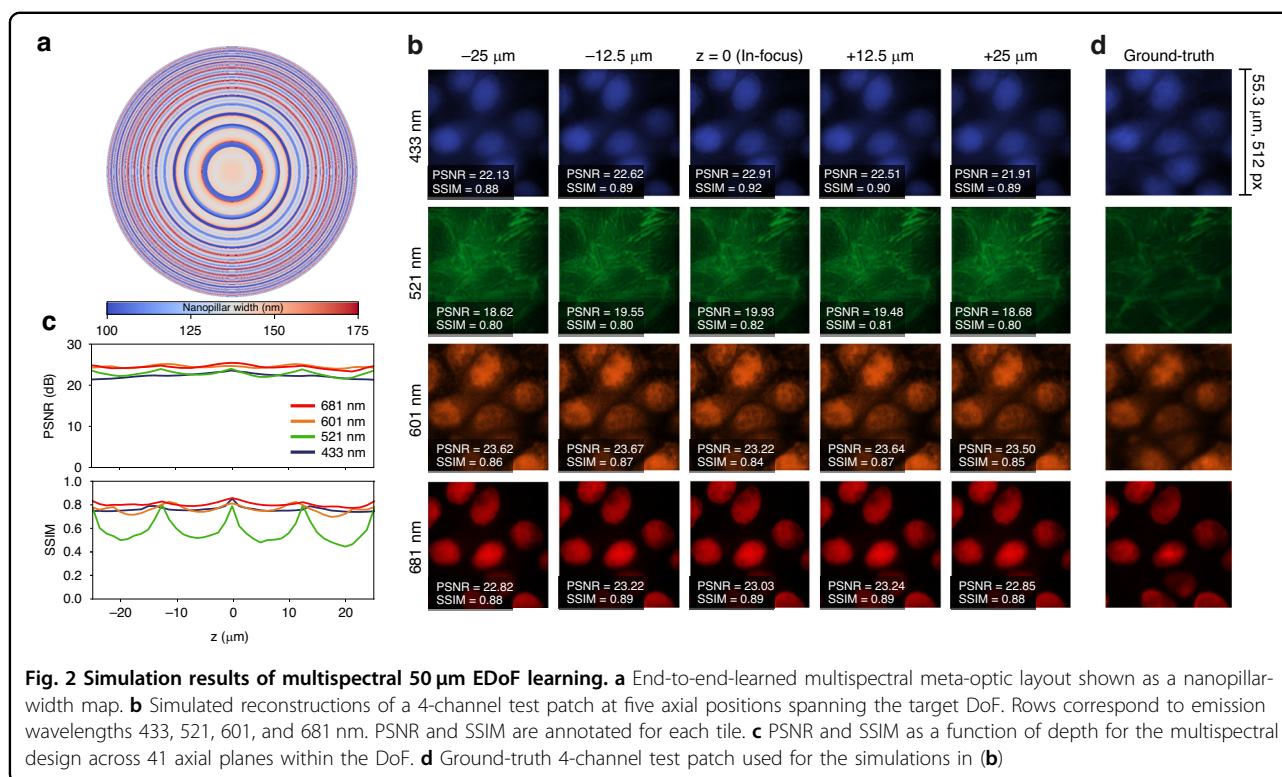
$$\Delta s = \frac{1}{\alpha} \Delta s_{\text{crit}}, \alpha = \frac{SBP_{\text{total}}}{SBP(\lambda_{\text{min}})} \quad (4)$$

With $\alpha = 4$, the resulting Δs oversamples the critical grid and provides a more than threefold margin in space bandwidth relative to the SBP requirement at the shortest wavelength, which reduces numerical error and improves PSF accuracy and reconstruction quality across wavelengths. Table 1 in Supplementary, reports the effect of the oversampling factor α on final performance.

Using these constraints, we trained configurations targeting EDoF ranges of 25, 50, 75 μm . Table 1 summarizes the Δs and ψ_{max} for each target training depth. To balance accuracy and computational efficiency, we set $K = 5$ training depths for the 25 μm and 50 μm targets, and $K = 7$ for 75 μm . Increasing K beyond these values did not improve the performance. With the multispectral meta-optic, PSFs vary smoothly with depth and differ across wavelengths, so a few well-spaced planes already cover the intermediate depths. Denser sampling adds redundant targets, and small wavelength-dependent differences increase the gradient noise without improving accuracy.

For all configurations, we implemented the network in PyTorch with two Adam optimizers: one for the optical parameters (learning rate 10^{-2}) and one for the D-CNN (learning rate 10^{-4} , weight decay 10^{-3}). Training and evaluation were performed on the same mixed dataset described in Section 4.3, using an NVIDIA Tesla V100 SXM2 with 32 GB memory. On average, each model required 44 hours and 36 minutes of training time. After training, during inference the D-CNN processes a four-channel 512×512 patch in 7.53 ms on average, corresponding to 132.7 patches per second on our hardware.





For quantitative evaluation, we assessed performance on 47 multispectral test images from our mixed dataset, across 41 uniformly spaced depth planes in diopters to cover the full axial range. Each four-channel image was assigned to a discrete depth, and metrics were aggregated over all depths and channels. Table 1 summarizes results for the target EDoF ranges using peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM). As expected, there is a gradual decrease in PSNR and SSIM as the targeted DoF increases from 25 to 75 μm , reflecting the increased difficulty of reconstructing scenes under stronger defocus.

For a target DoF of 50 μm , Fig. 2a displays the end-to-end-optimized meta-optic rendered as a nanopillar-width map. Representative reconstructions across wavelengths and axial positions are shown in Fig. 2b, where the ground-truth GT_λ is identical for all axial positions and is shown in Fig. 2d. Figure 2c reports the corresponding PSNR and SSIM versus depth for each spectral channel. The PSNR remains largely invariant with depth and exhibits minimal inter-channel variation, demonstrating stable pixel-level fidelity across the extended focal range. SSIM follows a similar trend for the 433, 601, and 681 nm channels; however, the 521 nm channel shows a decline outside the trained depth interval. This behavior likely reflects channel specific structure in the MDCK dataset, which consists of four channel fluorescence images of Madin Darby canine kidney epithelial cells labeled with

fluorophores emitting at 433, 521, 601, and 681 nm and acquired on a Nikon ECLIPSE Ti2 microscope, where the 521 nm channel emphasizes finer, lower contrast features that are intrinsically more challenging to reconstruct (see “Methods” and “Sample preparation” for details).

For comparison, earlier EDoF studies typically use small NA objectives and a restricted depth range, which leads to much smaller values of the defocus coefficient ψ_{max} , defined in Eq. (3). In contrast, our configuration uses a high NA and substantially larger target depth ranges, resulting in significantly higher ψ_{max} values. For reference, reported values are $\psi_{\text{max}} \approx 81.99$ for E2E-BPF²⁹, $\psi_{\text{max}} \approx 24.63$ for DeepDOF-SE³⁰, and $\psi_{\text{max}} \approx 24.59$ for Zhang et al.³¹. In our study, ψ_{max} ranges from 110 to 333 for the 25 μm , 50 μm , and 75 μm target DoFs, respectively. This indicates that our optimization operates under defocus blur conditions at least three times stronger than those considered in prior works. Moreover, while several of the compared designs are limited to a single wavelength, our framework is jointly trained and evaluated in a multispectral setting, further increasing the complexity of the optimization.

In addition to the multispectral design, we trained a single-wavelength meta-optic with a single-channel reconstruction network targeted at the 601 nm emission band. This model follows the UNet configuration from our recent work³² and omits the spectral Fourier domain attention block, since only a single channel is present and

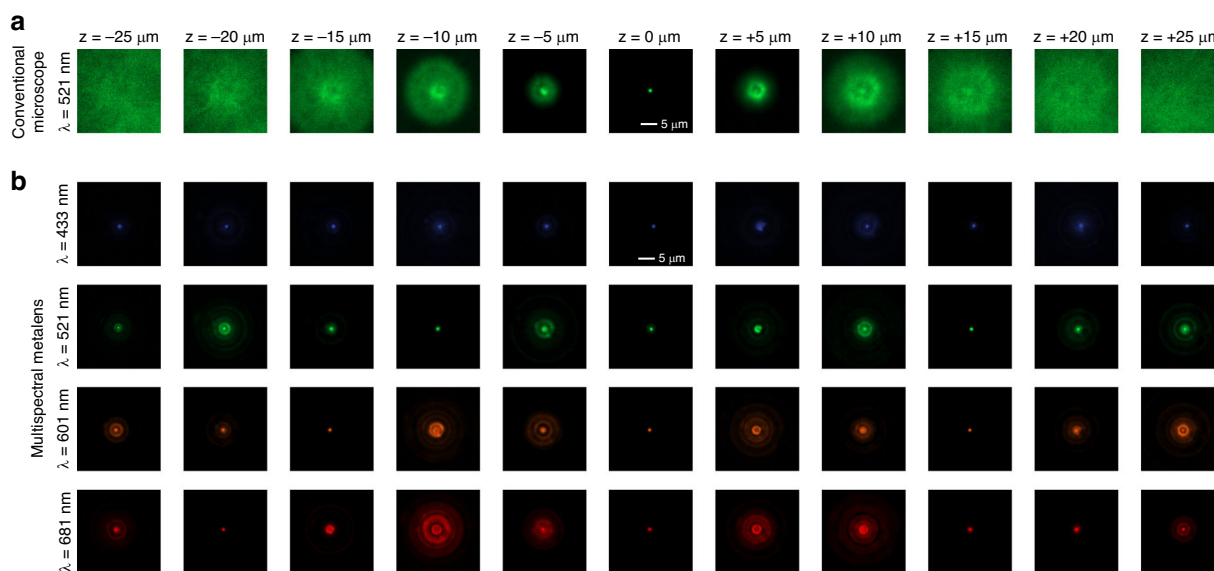


Fig. 3 Experimental PSFs across 50 μm target DOF. PSFs measured from 1 μm fluorescent beads for (a) a conventional microscope, b our multispectral meta-optic. For each configuration, PSFs are presented at 11 discrete axial positions spanning a 50 μm target DOF ($-25 \mu\text{m}$ to $+25 \mu\text{m}$ relative to best focus) and are shown with identical intensity scaling. Plots report the lateral FWHM versus depth, highlighting the stability of our designs. Scale bars, 5 μm

no cross-channel reweighting is needed. As detailed in Section 2 in the Supplementary, the single wavelength model achieves the highest PSNR and SSIM at its design wavelength of 601 nm, while performance at the other emission bands is lower. In contrast, the multispectral design sacrifices a small amount of peak performance at 601 nm and yields PSNR and SSIM that remain more uniform across wavelengths and depth, which is advantageous for quantitative multi-label fluorescence imaging. Together, these comparisons indicate that MANTIS learns a jointly optimized multispectral EDoF response with stable reconstruction quality across both wavelength and depth.

Experimental results

To validate MANTIS, we fabricated the multispectral meta-optic designed for a 50 μm target EDoF and integrated it into the 4f relay described above. As reference configurations, we use a conventional wide-field microscope and a single wavelength meta-optic optimized at 601 nm. We evaluate these configurations via PSF measurements and reconstructions of biological samples, as detailed below. Unless stated otherwise, wide-field and meta-optic images are acquired in parallel using two arms of the same microscope, providing a frame matched baseline for each FoV (see Supplementary Fig. 1 and Supplementary Section 1).

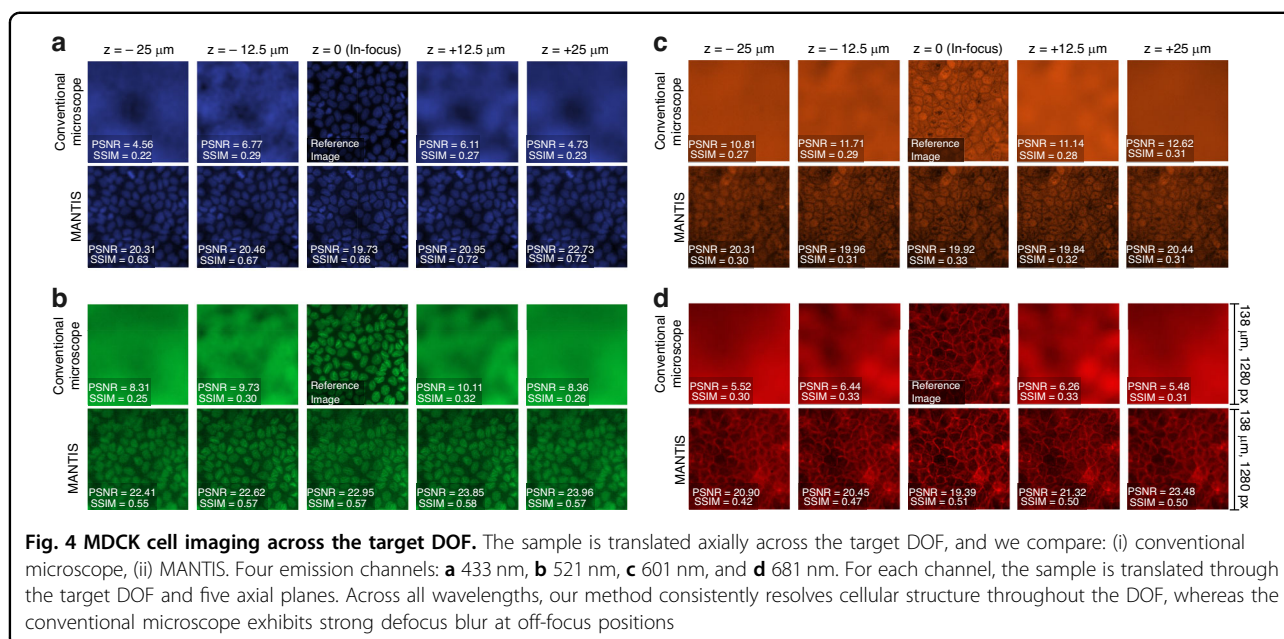
PSF characterization

We experimentally characterized the fabricated meta-optic and the reference wide-field microscope by

measuring PSFs from 1 μm fluorescent beads embedded in a polyacrylamide hydrogel under LED excitation at each emission wavelength. The sample was axially scanned over a 50 μm range with 500 nm step size using a piezo stage, and fluorescence images were captured at each axial position to obtain a three-dimensional PSF stack. Experimental PSFs for the conventional microscope and for the multispectral meta-optic are shown in Fig. 3. Compared with the wide-field microscope, its PSFs vary much less with defocus and with wavelength, indicating reduced axial sensitivity and improved lateral invariance across the scan. For comparison, simulated PSFs computed under the same conditions are shown in Supplementary Fig. 6 and agree well with the measurements, supporting that the meta-optic encodes the intended spectral phase profile. For completeness, we applied the same protocol to the learned single wavelength meta-optic and report the corresponding measured PSFs at each emission band in Supplementary Fig. 3.

To characterize PSF width and its variation with depth for the multispectral meta-optic, we fitted a two-dimensional Gaussian to the measured PSFs at each wavelength and axial plane and extracted the lateral full width at half maximum (FWHM). We displayed these fits as horizontal and vertical line profiles with the Gaussian curves overlaid, and the corresponding FWHM values indicated on each panel. Plots for all four emission bands and all depth planes are provided in Supplementary Section 3. The resulting FWHM curves are consistent with achromatic EDoF behavior of the learned multispectral





meta-optic and show only modest broadening over depth and wavelength.

We note that metasurface fabrication can introduce deviations from the designed nanopillar geometry, which affect the realized wavelength-dependent phase profile and can shift the measured PSF relative to the nominal design. This introduces a model mismatch that may reduce reconstruction fidelity when deviations are sufficiently large. In our framework, robustness can be improved by calibrating the forward model using experimentally measured PSFs of the fabricated meta-optic and, if needed, fine tuning the reconstruction network on calibration data. In the present experiments, this additional fine tuning was not required, since the measured PSFs closely followed the expected depth and spectral trends. A quantitative analysis of fabrication induced geometry perturbations for learned meta-optics and their impact on imaging performance is provided in our prior work³².

Reconstruction of biological samples

We evaluated MANTIS on three biological specimen types: MDCK epithelial monolayers, three-dimensional cultured MDCK II spheroids ($\approx 50 \mu\text{m}$ thick), and a mouse kidney cryosection (FluoCells Prepared Slide 3, F24630, $16 \mu\text{m}$ thick). Detailed preparation and staining protocols are provided in “Methods”, Section “Sample preparation”.

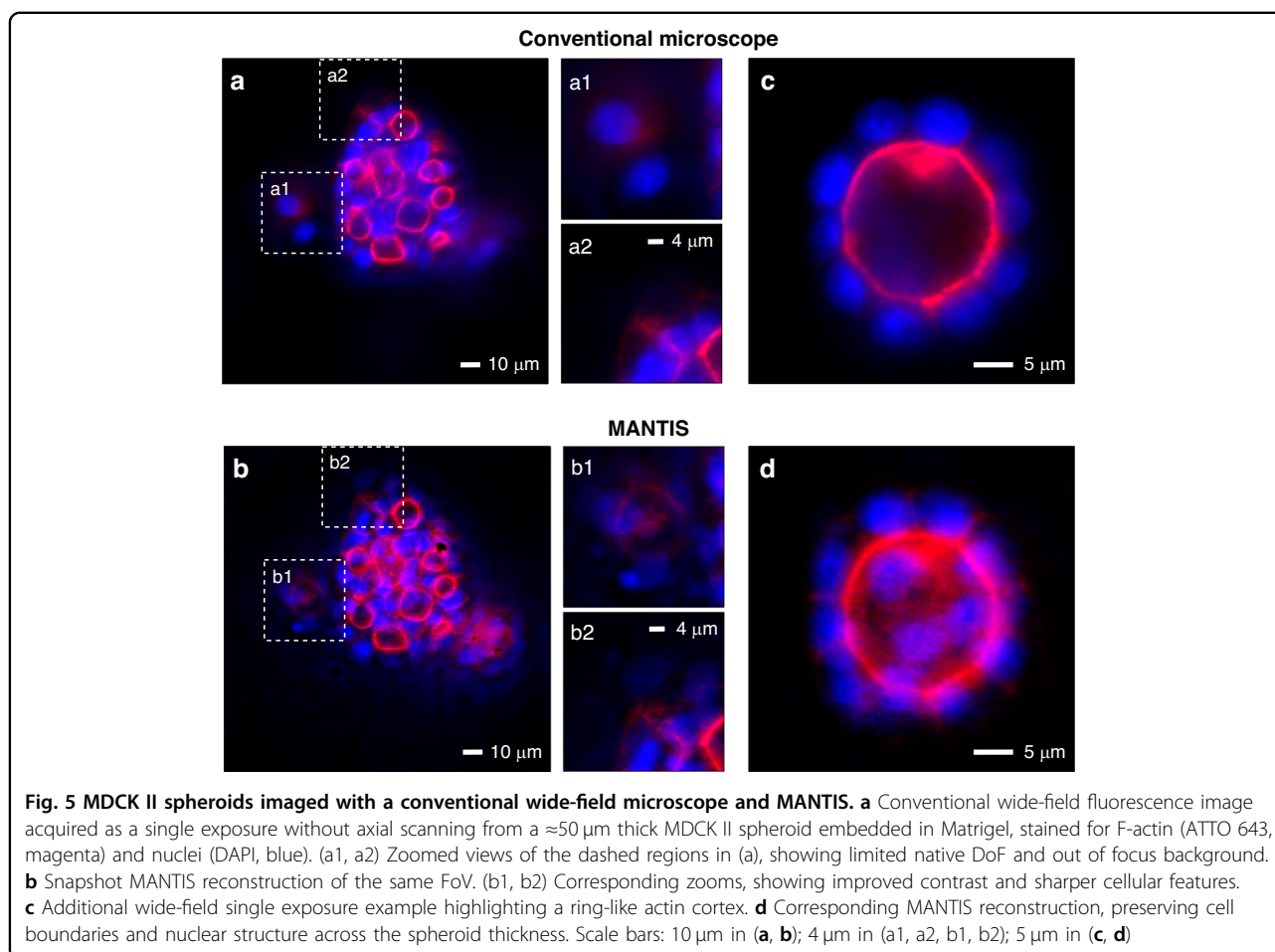
We first imaged MDCK epithelial monolayers stained with the same four fluorescence markers used in the dataset, namely DAPI for nuclei, Atto Fluor 640 for F-actin, Alexa Fluor 568 for vimentin, and Alexa Fluor 488 for lamin A, thereby capturing both nuclear and cytoskeletal structures. For each FoV, we acquired paired

measurements on a standard wide-field microscope and on the microscope incorporating the learned multi-spectral meta-optic. In both cases, the sample was translated axially through the target DoF and imaged at five discrete planes. Only the meta-optic measurements were processed through the end-to-end trained D-CNN, while the wide-field images are shown as acquired to provide a baseline reference.

Figure 4 compares MDCK images across depth and wavelength. In the wide-field images, cellular morphology rapidly deteriorates with defocus, and blur dominates the contrast. Using our approach, fine cellular features remain discernible throughout the targeted DoF. For quantitative assessment, we took the in-focus wide-field image as the reference when computing PSNR and SSIM as functions of depth. Within each spectral channel, both metrics exhibit only modest variation with axial position. These experimental trends align with our simulated performance across depth and wavelength and are consistent with the low chromatic aberration of the meta-optic. During inference, on an NVIDIA Tesla V100 SXM2 with 32 GB memory, D-CNN processed 1280×1280 pixel patches in 46.97 ms on average (21.3 patches per second).

Using these experimental results, we further assessed the role of co-design by quantifying the contributions of the meta-optic and the D-CNN in Supplementary Section 4. We compared three configurations: the full system with both meta-optic and D-CNN, a system with the meta-optic removed, and a system with the D-CNN removed, all based on the same D-CNN architecture. These ablation experiments show that neither component on its own matches the performance of MANTIS in our experiments,





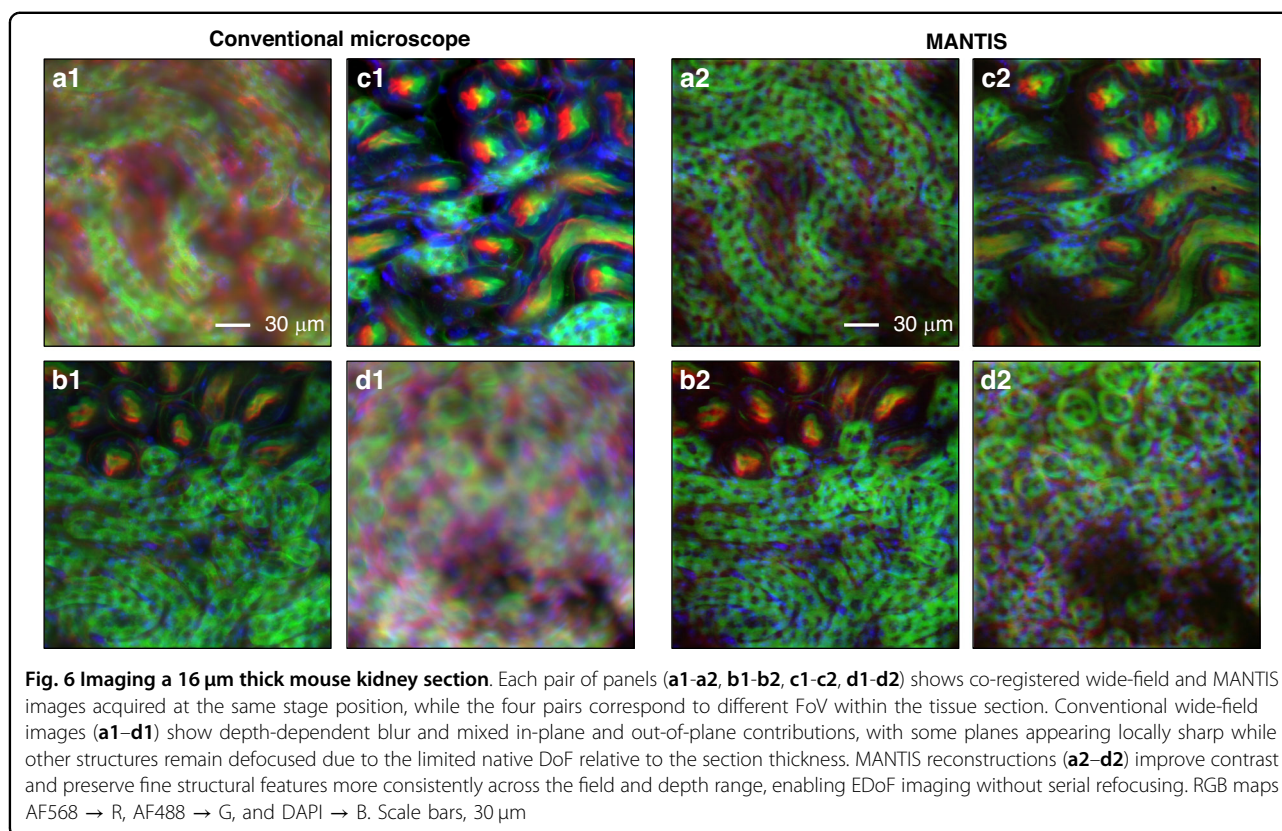
highlighting the importance of joint optimization of optics and reconstruction.

For completeness, we repeated the experiment with the single wavelength learned meta-optic optimized at 601 nm and reconstructed those measurements with the 1-channel D-CNN. The PSF measurements were well-formed across wavelengths, yet the reconstructions show strong chromatic dependence. Image quality is strongly wavelength dependent, with PSNR and SSIM varying substantially across channels. Results are provided in Section 2 in the Supplementary.

To further assess the DoF performance of MANTIS, we imaged 3D-cultured MDCK II cells ($\approx 50\ \mu\text{m}$ thick), which form spheroids with hollow lumens rather than a monolayer. The samples were labeled with ATTO 643-phalloidin (F-actin) and DAPI (nuclei). Figure 5 compares conventional wide-field images and MANTIS reconstructions for two representative fields of view. In the conventional microscopy images (Fig. 5a, c), only a narrow axial region of each spheroid lies within the native DoF, so out-of-plane structures appear blurred and low contrast. The actin cortex is diffused, and neighboring cells tend to merge, and in Fig. 5c, the ring-like cortex

around the lumen is only partly in focus, and several nuclei are dimmed or smeared. With MANTIS (Fig. 5b, d), we obtain that the spheroids remain sharper. This is most evident in the zoomed regions marked by the white dashed boxes (Fig. 5a1, a2 vs. Fig. 5b1, b2): MANTIS preserves cell boundaries and clearly resolves individual DAPI-labeled nuclei throughout the 3D structure, whereas the conventional images show smeared cortices and merged cell profiles. As an additional reference, we acquired conventional wide-field axial z stacks and performed Richardson-Lucy deconvolution, reporting maximum intensity projections of the deconvolved volumes for visualization (Supplementary Section 10).

Furthermore, to evaluate performance on a specimen type not used during training and to assess the generalizability of MANTIS, we imaged a $16\ \mu\text{m}$ mouse kidney section (FluoCells Prepared Slide 3, F24630) stained with three fluorescent labels (nuclei with DAPI, cell membranes with WGA conjugated Alexa Fluor 488, F-actin with Alexa Fluor 568). The cellular length scales are comparable to those in MDCK epithelial samples, but kidney tissue shows a more heterogeneous organization, with renal tubules and glomerular regions. DAPI labels



nuclei, AF488 delineates renal tubules, and AF568 highlights F-actin, providing a three-channel, multispectral test in tissue. For each FoV, we acquired two datasets: a conventional wide-field image and a measurement using a microscope equipped with the learned multispectral meta-optic. Only the meta-optic measurements were processed with the end-to-end trained D-CNN; the wide-field images are shown without post-processing and serve as the baseline. Both systems imaged the same tissue regions at identical out-of-focus planes, revealing that the meta-optic system, combined with the CNN, restores the fine tubular and cytoskeletal structure that remains unresolved in conventional wide-field microscopy.

Figure 6 compares conventional wide-field imaging with MANTIS at matched axial offsets. For each pair, the wide-field image and the corresponding MANTIS measurement are acquired at the same stage position, with axial offsets defined relative to a common reference plane. We define $z = 0 \mu\text{m}$ as the nominal mid-plane of the 16 μm -thick section. To establish this reference, we first located the coverslip facing surface by translating the stage toward the objective until the section surface came into focus, then moved 8 μm into the specimen and re-optimized the wide-field focus. This mid-plane focus position was used as $z = 0 \mu\text{m}$, and all reported z_i values are stage offsets relative to this plane. In Fig. 6, the pairs (a1, a2) to (d1, d2) were acquired at $z_i = -7 \mu\text{m}$, $-3.5 \mu\text{m}$, $+3.5 \mu\text{m}$, and

$+7 \mu\text{m}$, respectively, and correspond to different FoV within the section.

Across these axial offsets, the wide-field images (a1–d1) exhibit increasing defocus blur and strong out-of-focus background. Even when some structures are near best focus, their visibility is reduced by fluorescence from the rest of the section. In contrast, the MANTIS reconstructions (a2 to d2) retain higher contrast and sharper boundaries over the same offsets, consistent with EDoF behavior, particularly in dense regions such as glomeruli and surrounding tubules.

Finally, we emphasize that MANTIS provides an EDoF reconstruction rather than optical sectioning and does not refocus to distinct depth planes within the tissue. The purpose of varying z_i in Fig. 6 is to evaluate robustness to axial displacement relative to the reference plane. The four examples correspond to different FoV, so differences in the high contrast structures reflect specimen variability and local optical heterogeneity across the section.

Discussion

We introduce MANTIS, a computational fluorescence microscopy system that jointly learns a multispectral meta-optic and a D-CNN through end-to-end training. In the optical path, a TiO_2 nanopillar meta-optic is placed at the conjugate Fourier plane of a 4f relay, where it applies wavelength-dependent pupil phase coding to reshape the

system PSF across depth and wavelength. The recorded sensor image is then processed by a modified UNet to reconstruct the object. In a conventional high NA microscope, defocus introduces a second-order wavefront error that imposes a depth-dependent quadratic phase at the pupil, causing the PSF to vary rapidly with axial position. In parallel, chromatic dispersion produces wavelength-dependent focal shifts and aberration changes, so the optimal focus plane and blur characteristics differ across spectral channels. In MANTIS, the learned phase profile compensates these effects by shaping the pupil such that the modulation transfer function (MTF) varies more slowly with defocus over the target range. This preserves informative measurements under strong defocus across wavelengths and improves the conditioning of the inverse problem. The full system is trained using a physics-based differentiable forward model and a multi-term loss, enabling optical encoding and computational decoding to co-optimize for depth robust and spectrally consistent reconstructions.

MANTIS delivers EDoF performance over targeted ranges of 25–75 μm in simulation, and we experimentally validate the 50 μm configuration. Across depth and wavelength, PSNR varies only modestly with axial position and cross-spectral changes are minimal, while SSIM trends track the structural content of each channel. Measured PSFs confirm reduced axial sensitivity and strong chromatic stability relative to a conventional microscope. Reconstructions of biological specimens further demonstrate that fine cellular structure remains visible and sharp throughout the extended focal range.

We expect MANTIS to generalize to new specimens if the microscope configuration is fixed and the fluorophore emission falls within the same four spectral bands used during training. Consistent with this, we imaged a mouse kidney cryosection, an unseen tissue type with different morphology from the MDCK data and obtained robust reconstructions. Generalization may degrade when imaging conditions deviate substantially from training, for example, in highly scattering or strongly autofluorescent specimens, under markedly different spatial statistics, or when emission lies outside the trained bands. For new specimen types within similar emission bands, fine-tuning the reconstruction network with a calibration dataset acquired through the meta-optic can recover performance. By contrast, changes in fluorophore labeling or the spectral configuration alter the emission spectra and thereby change the wavelength-dependent PSFs encoded by the meta-optic. In this case, end-to-end retraining may be required to match the new spectral regime.

While our results quantify performance under high NA and large defocus coefficients, it is important to recognize that prior EDoF methods are often optimized for different points on the tradeoff between DoF range, lateral

resolution, and spectral bandwidth. For example, some approaches prioritize near diffraction-limited and depth uniform PSFs over a shorter EDoF, whereas others target a larger EDoF but reduced lateral resolution, and many are optimized for a single wavelength. In contrast, MANTIS is designed to operate under substantially stronger defocus while jointly satisfying multispectral constraints. This choice inherently trades off peak lateral sharpness at a single plane for improved reconstruction stability across depth and wavelength. We evaluate this primarily using image-level performance, reporting PSNR and SSIM versus depth, together with the corresponding wide-field baseline curves under the same reference definition. Additionally, we report the FWHM measured from 1 μm beads as an indicator of the raw optical PSF and its depth dependence. Because MANTIS includes computational reconstruction, this bead-based FWHM does not directly represent the final lateral resolution after postprocessing.

To assess the benefit of multispectral learning, we also evaluated a single-wavelength model trained at 601 nm with a one-channel reconstruction network. In simulations, this configuration provides higher spatial fidelity near its design wavelength, but experimentally, the reconstructions show strong wavelength dependence, with image quality varying across channels. In contrast, the multispectral model learns jointly across channels and maintains uniform performance, enabling robust multispectral fluorescence imaging.

In conclusion, MANTIS combines end-to-end multispectral co-optimization of the meta-optic and reconstruction network with an MTF-guided frequency-domain initialization and a well-defined Fourier-plane sampling strategy. Unlike earlier EDoF approaches that trade off either NA or sample complexity, MANTIS enables high NA fluorescence imaging of thick, multi-label biological specimens in a single shot, providing a practical route to high-resolution multispectral EDoF imaging. Future work will incorporate measured system aberrations into the design, extend the framework to broader spectral coverage and larger EDoF ranges, and advance the system toward three-dimensional reconstruction of volumetric samples.

Methods

Microscope model

The microscope model simulates the forward image formation process by receiving high-resolution fluorescence image patches GT_λ as input, where each patch is a fixed-size spatial crop extracted from a full FoV image, together with a depth z , and generating the corresponding sensor image $SI_{z,\lambda}$. At each iteration, a depth and wavelength-dependent $PSF_{z,\lambda}$ is computed based on wave optics theory. The sensor image is then formed by



convolving the ground-truth image with the simulated PSF:

$$SI_{z,\lambda} = GT_{\lambda} * PSF_{z,\lambda} + \eta \quad (5)$$

where η denotes sensor noise. A detailed derivation of the forward imaging model and PSF formulation is provided in Section 5 in the Supplementary.

To compute each PSF, the model constructs a spatially varying phase profile $\phi_{z,\lambda}$, which consists of a defocus phase term $\phi_{z,\lambda}^{DF}$ and a wavelength-dependent phase modulation term ϕ_{λ}^M . The modulation term ϕ_{λ}^M is determined by the underlying geometry of the meta-optic, which is learned during training, and is computed using a differentiable spectral lookup function implemented in the meta-optics simulator part, as illustrated in Fig. 1b:

$$\phi_{\lambda}^M = \text{Interp}(U_{\gamma}, U_{\phi_{\lambda}}, \gamma) \quad (6)$$

where γ denotes the meta-optic radius vector, U_{γ} is the set of discrete geometric widths, and $U_{\phi_{\lambda}}$ contains the corresponding wavelength-dependent phase delays obtained from electromagnetic simulations of meta-optic unit cells. Details of the unit cell library construction and the corresponding electromagnetic simulation results are provided in Section 6 in the Supplementary.

Before end-to-end training, we pre-optimize the meta-optic alone for 30 epochs using an EDoF frequency-domain objective. From the simulated stack $PSF_{z,\lambda}$, we compute the MTF at each depth z and wavelength λ , normalize by the DC term, and form a DC-normalized radial profile over normalized spatial frequency $\rho \in [0, 1]$, denoted $M_{\lambda}(\rho, z)$. We compare this profile to a target passband $T(\rho)$ that is unity up to a cutoff ρ_{cut} . For each wavelength, depth errors are aggregated with a log-sum-exp operator as

$$\mathcal{L}_{EDoF,\lambda} = \frac{1}{\beta} \log \sum_z \exp(\alpha \|M_{\lambda}(\rho, z) - T(\rho)\|_1) \quad (7)$$

Here $\beta > 0$ controls the aggregation across depths: A large β approaches $\max_z$ (most defocused plane dominates), whereas a small β approaches the average over depths. In practice, we use a large β so that “soft-max” pooling emphasizes the worst-case depth and drives the design toward meta-optic parameters that maintain image quality at the largest defocus. The result is a single depth-robust score $\mathcal{L}_{EDoF,\lambda}$ for each wavelength. To balance performance across wavelengths, we then aggregate the per-channel scores as follows:

$$\mathcal{L}_{EDoF} = \lambda_{focus} \mathcal{L}_{EDoF,\lambda^*} + (1 - \lambda_{focus}) \frac{1}{\eta} \log \sum_{\lambda} \exp(\eta, \mathcal{L}_{EDoF,\lambda}) \quad (8)$$

where $\lambda^* = \text{argmax}(\mathcal{L}_{EDoF,\lambda})$ is the channel that currently limits and $\eta > 0$ controlling the aggregation across channels (large η approaches the maximum over λ , small η approaches the mean). The coefficient $\lambda_{focus} \in [0, 1]$ places extra weight on the limiting channel. This pre-optimization yields a multispectral EDoF seed that produces sensor images with manageable blur and stabilizes the subsequent end-to-end training. We also tested a commonly used cubic-phase initialization²⁵, which produced extremely blurred sensor images under the large defocus and reduced the effective signal for the learned D-CNN. In contrast, a flat π -phase initialization, combined with the EDoF-MTF pre-optimization, provides a physically meaningful starting point that evenly distributes contrast across depth and wavelength. By comparison, for a conventional wide-field microscope equipped only with a learned post-processing network (and no meta-optic), the situation is analogous but more constrained: native defocus blur dominates the measurement, leaving the sensor data severely out of focus, and the inverse problem poorly conditioned.

During training, the reconstruction loss is back-propagated through the microscope model, enabling gradient-based optimization of the meta-optic design. Specifically, loss gradients with respect to the PSF, $\partial \mathcal{L} / \partial PSF_{z,\lambda}$, are propagated through the wave optics model to compute the derivative with respect to the nanopillar geometry, $\partial \mathcal{L} / \partial \gamma$. In this parameterization, an effective edge width parameterizes the meta-optic radius vector γ :

$$\gamma = \gamma_{init} + \Delta \gamma \quad (9)$$

where γ_{init} is a fixed initialization based on the frequency-domain EDoF objective, and $\Delta \gamma$ is a learnable perturbation. The full 2-D radius map is formed by azimuthally replicating radius vector K times,

$$\Gamma_{full} = \text{Rot}_K(\gamma) \quad (10)$$

where $\text{Rot}_K(\cdot)$ is a fixed linear replication operator with no learnable parameters. A differentiable edge-to-phase lookup, $\text{Interp}(U_w, U_{\phi_{\lambda}}, \gamma)$, then yields the per-wavelength phase ϕ_{λ} . By the chain rule, $\partial \mathcal{L} / \partial \gamma$ is obtained from $\partial \mathcal{L} / \partial PSF_{z,\lambda}$ through the wave-optics model and the lookup, updating the learned radius vector and producing an end-to-end optimized meta-optics. Detailed gradient derivations are provided in Section 7 in the Supplementary.

D-CNN

The proposed deblurring module takes the sensor image $SI_{z,\lambda}$ and predicts a four-channel reconstructed image aligned to the high-resolution ground-truth



fluorescence images GT_λ . We adopt a three-level encoder-decoder UNet with skip connections³³, inspired by recent advances in biomedical image restoration^{17,29–31}. Before the encoder, each spectral channel is normalized using InstanceNorm to decouple features from channel-specific intensity and contrast variations that naturally occur in multispectral data due to different fluorophores or sensor responses. The channels are then linearly mixed by a 1×1 convolution initialized to identity. This 1×1 convolution provides the network with a flexible, learnable starting point for spectral feature extraction, allowing it to automatically discover the optimal linear combination of the normalized channels before spatial processing begins. A Fourier-domain channel-attention block then computes per-channel gates from the magnitude spectra of the feature channels and uses them to reweight the features, providing wavelength-aware balancing at minimal cost.

Each encoder block consists of a 2×2 max-pooling operation followed by two 3×3 convolutional layers, each with batch normalization and ReLU activation. The feature depth increases progressively from 32 to 128 across the encoder. For the decoder, bilinear upsampling is applied at each stage, followed by the same double-convolution structure to recover spatial resolution while reducing feature depth. A final 1×1 convolution is used to project the output to four channels, corresponding to the spectral components. The network learns a residual image through a hyperbolic tangent (tanh) activation function, which is added to the input $SI_{z,\lambda}$ to reconstruct the deblurred output. The deblurred output is passed through a Clamp layer to keep the pixel values between 0 and 1. This step helps avoid unrealistic results and maintains the training stability.

To ensure memory efficiency, the network operates on 512×512 pixel patches corresponding to specific depth and wavelength combinations. Each input patch is initially convolved with the corresponding PSF and subsequently cropped to 512×512 pixels to suppress boundary artifacts introduced by the convolution operation.

End-to-end learning

The end-to-end network is trained on fluorescence images of MDCK epithelial cells (Fig. 2) acquired with a Nikon ECLIPSE Ti2 inverted microscope and a CFI Plan Apo 60 $\times$ /1.4 NA oil-immersion objective (see Section “Sample preparation” for full preparation and staining details). To introduce spectral variability and prevent overfitting, we expanded the dataset with natural images from the Hyperspectral Image Dataset³⁴. The final mixed dataset contains 423 MDCK fluorescence images and 37 hyperspectral natural images, corresponding to approximately 92 percent and 8 percent of the dataset, respectively. The hyperspectral images were converted to four-channel inputs by selecting bands whose center

wavelengths are closest to our four fluorescence emission channels. In total, the mixed dataset comprises 460 images of size 512×512 pixels, split into training, validation, and test sets at 80%, 10%, and 10%. This mixed dataset encourages the network to learn spatial and spectral features across varying scales, thereby improving its robustness and generalization. During training, data augmentation is performed through random rotations, flips, and brightness variations. At each iteration of the training and validation stages, a randomly selected in-focus ground-truth image patch is assigned to one of K discrete depth positions $z_i \in (z_1, \dots, z_K)$, uniformly distributed in diopters across the target DoF range. The assigned depth z selects the corresponding $PSF_{z,\lambda}$ used to synthesize $SI_{z,\lambda}$. The ground-truth is depth independent, that is $GT_{z,\lambda} = GT_\lambda$ for all z . During each forward pass, output images are computed for each selected depth and wavelength, and the loss is accumulated and averaged across all combinations.

We use a fixed multi-term loss to capture fidelity, structure, edges, and coarse content. For each (z, λ) patch:

$$\mathcal{L} = \frac{1}{KC} \sum_{z,\lambda} [w_{\text{RMSE}} \text{RMSE}(OI_{z,\lambda}, GT_{z,\lambda}) + w_{\text{SSIM}} (1 - \text{SSIM}(OI_{z,\lambda}, GT_{z,\lambda})) + w_{\text{grad}} \ell_{\text{grad}}(OI_{z,\lambda}, GT_{z,\lambda}) + w_{\text{LP}} \text{MSE}(\phi_{\text{LP},k}(OI_{z,\lambda}), \phi_{\text{LP},k}(GT_{z,\lambda}))] \quad (11)$$

where K and C are the numbers of depth planes and spectral channels. ℓ_{grad} matches horizontal and vertical finite differences to preserve edges, and $\phi_{\text{LP},k}$ denotes a fixed low-pass filtered representation used in an MSE term to align large-scale structure and intensity balance. During learning, RMSE enforces pixel level fidelity and provides the dominant optimization signal under our weighting, while SSIM, gradient matching, and low-pass MSE act as complementary regularizers.

We optimize the end-to-end network using two Adam optimizers³⁵: one for the optical parameters (no weight decay) and one for the D-CNN parameters with $L2$ regularization. The total training objective is:

$$\mathcal{J} = \mathcal{L} + \beta \|\theta_{\text{deblur}}\|_2^2 \quad (12)$$

while the optical parameters are not regularized.

Since the input undergoes varying levels of blur across depth and wavelength, the network naturally adapts the meta-optic to produce PSFs that are less sensitive to defocus and spectral changes. As a result, no additional loss term is needed to enforce PSF similarity within the targeted depth range explicitly.

Sample preparation

We evaluated three specimen types. (i) MDCK II epithelial cells cultured on #1.5H coverslips and immunolabeled for lamin A/C (Alexa Fluor 488) and vimentin



(Alexa Fluor 568); F-actin (Atto 640) was labeled with phalloidin and nuclei with DAPI. (ii) A fixed mouse kidney cryosection (16 μm ; FluoCells Prepared Slide #3, F24630) labeled with DAPI, Alexa Fluor 488, and Alexa Fluor 568. (iii) Polyacrylamide hydrogels (11 kPa) with embedded 1 μm fluorescent beads used to measure the PSF, assess lateral resolution (FWHM). Full protocols, reagent sources, catalog numbers, dilutions, and timings are provided below.

Madin-Darby canine kidney II (MDCK II) epithelial cells were maintained in media consisting of Minimum Essential Medium (MEM) with GlutaMAX and Earle's salts (#41090028, Gibco, USA), 10 % FBS (#10500064, Gibco) and 83 U/ml penisillin/streptomycin (#15140122, Gibco) at 37 °C in 5% CO₂. The cells were passaged using Trypsin-EDTA (0.25 %, #25200056, Gibco) and plated at approximately 6×10^4 cells/cm² on 22 × 22 mm glass coverslips (No. 1.5H, Marienfeld, Germany) in 6-well plates (#3506, Corning, USA) and cultured for 5 days. All the following steps were conducted at room temperature. Cells were fixed in 4% paraformaldehyde (v/v, #157–8, Electron Microscopy Sciences, USA) in 1x PBS for 10 min and washed 3 times in PBS for 5 min. Permeabilization of the cells was done in 0.5% Triton-X-100 (v/v, #T8787, Sigma-Aldrich, USA) in 0.5% BSA in PBS for 10 min, followed by blocking in 3% bovine serum albumin (BSA) (w/v, #P06-139210, PAN-Biotech, Germany) in PBS for 1 h. Primary antibodies lamin A+C [131C3] (#ab8984, abcam, UK) in 1:300 dilution and vimentin (#ab92547, abcam) in 1:200 dilution in 3% (w/v) BSA in PBS were incubated for 1 h. Samples were washed three times for 10 min, first in the permeabilization buffer, then in PBS, and finally again in the permeabilization buffer. Secondary antibodies goat anti-mouse Alexa Fluor 488 (#A-11029, Invitrogen, USA) and goat anti-rabbit Alexa Fluor 568 (#A11011, Invitrogen, USA) in 1:200 dilution in 3% BSA (w/v) in PBS were incubated for 1 h in dark. Cellular actin was labeled with ATTO 643 Phalloidin (#AD643-81, ATTO-TEC, Germany) in 1:100 dilution with the secondary antibodies. The samples were washed 2 times in PBS for 10 min and stained for DAPI in 1:1000 dilution in ddH₂O. Finally, the samples were washed twice in PBS and mounted in Prolong diamond antifade mountant (#P36961, Invitrogen). Samples were kept overnight in room temperature and then stored at 4 °C. Samples were imaged on a Nikon ECLIPSE Ti2 with a CFI Plan Apo 60 × /1.4\$, NA oil-immersion objective.

In addition, MDCK II cells were cultured in a 3D environment, where cells form spheroids with empty lumens instead of a monolayer. Cells were cultured and passaged as described above. Matrigel (#354277, Corning) was thawed overnight on ice and kept below 4 °C prior to gelation. Cells were mixed with Matrigel, and 40 μl

droplets were pipetted into the wells of a chambered coverslip suitable for microscopy (#80807, Ibidi), containing approximately 8000 cells in each gel. The gels were polymerized for 15 min inside the cell culture incubator. 300 μl of cell culture media was added to each well, and the samples were cultured for 7 days, with media changes every 2–3 days. Cells were fixed in 4% paraformaldehyde with 0.01% Glutaraldehyde (#340855, Sigma-Aldrich) in 1× PBS solution pre-warmed to 37 °C for 20 min at 37 °C. The following steps were conducted at room temperature. The samples were washed 3 times in PBS for 5 min. Cells were permeabilized for 45 min and labeled with ATTO 643 Phalloidin and DAPI, as described above. The samples were stored at 4 °C in 1x PBS.

A commercially prepared slide was used (FluoCells Prepared Slide #3; F24630, Thermo Fisher Scientific/Invitrogen). The slide contains a 16 μm cryostat section of mouse kidney labeled with DAPI, WGA conjugated to Alexa Fluor 488, and Phalloin conjugated to Alexa Fluor 568. Slides were imaged as supplied, following the vendor's instructions; no additional processing was performed.

Polyacrylamide (PA) hydrogels with Young's modulus of $E = 11$ kPa were prepared following the protocol of³⁶. The hydrogels were prepared between two glass coverslips. Both the bottom coverslips (22 × 22 mm, No. 1.5H, #0107052, Marienfeld) and top coverslips ($d = 13$ mm, #631–0150, VWR, USA) were washed with 2% Helmanex III (#Z805939, Hellma, Germany) in ddH₂O solution for 30 min in sonicator (CPX3800H-E Branson, Branson, USA), rinsed with dH₂O and EtOH, and dried with compressed nitrogen. The bottom coverslips were treated with 0.3% v/v bind-silane (3-(trimethoxysilyl) propyl methacrylate (#M6514, Sigma-Aldrich) and 5% v/v glacial acetic acid solution in 95% ethanol to facilitate PA gel adhesion. Next, 10% acrylamide (stock 40%, #1610140, Bio-Rad Laboratories, USA) and 1% Bis-acrylamide (Bis) (stock 2%, #1610142, Bio-Rad Laboratories) were mixed with PBS and degassed in a vacuum. 0.04% v/v of fluorescent beads (1 μm tetraspecks, #T7282, Invitrogen) was added to the solution. To initiate PA polymerization, 0.2% v/v of TEMED (#1610800, Bio-Rad Laboratories) and 1% v/v of ammonium persulfate (APS) stock solution in ddH₂O (10% w/v, #A3678, Sigma-Aldrich) were mixed to the solution. Twenty six microliters droplet was pipetted on the bottom coverslip and the top coverslip was placed on top of the droplet to construct approximately $h = 200$ μm gels. The PA gels were polymerized for 15 min in room temperature and after polymerization were stored in PBS at 4 °C.

Meta-optics fabrication

The learned multispectral meta-optic design is fabricated using electron beam lithography and reactive ion



etching on a fused silica substrate coated with a high-refractive-index TiO_2 layer. Details of the fabrication steps are provided in in Section 8 in Supplementary. As a reference, the end-to-end learned single-wavelength meta-optic design was fabricated by the same process. Both meta-optics are integrated into the microscope at the Fourier plane of a $4f$ system via a relay configuration; a detailed schematic of the experimental setup is provided in Supplementary Fig. 1. This placement ensures that phase modulation occurs in the angular domain, enabling PSF shaping across both spatial and spectral dimensions.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 956770. This work was supported by the Academy of Finland (project no. 336357, PROFI 6—TAU Imaging Research Platform). The Jane and Aatos Erkkö Foundation supported this work through the Future Makers program (project "High-speed 3D Microscopy"). The authors thank Turku Salminen for assistance with SEM imaging, Linzhi Yu for help with the thick-sample experiments and discussions on the fabrication, and Charles Kervran for insightful comments and suggestions on the manuscript. We also acknowledge the Tampere Microscopy Center facilities at Tampere University for providing the necessary infrastructure. The authors thank Sanni Erämies regarding her contribution to imaging of MDCK samples, and acknowledge the Biocenter Finland (BF) and Tampere Imaging Facility (TIF) for the service.

Author details

¹Physics Unit, Faculty of Engineering and Natural Sciences, Tampere University, Tampere 33720, Finland. ²INRIA Rennes, Bretagne Atlantique Research Centre, Rennes 35042, France. ³Bio Unit, Faculty of Medicine and Health Technology, Tampere University, Tampere 33520, Finland. ⁴Signal Processing Research Centre, Tampere University, Tampere 33720, Finland. ⁵Department of Electrical Engineering, Photonic Integration Group, Eindhoven University of Technology, Eindhoven 5600 MB, The Netherlands

Author contributions

I.A.A. developed MANTIS, designed meta-optics and an end-to-end computational framework, implemented the reconstruction pipeline, performed simulations, and fabricated meta-optics. J.S. developed the microscopy setup and performed the experiments. S.K. and T.O.I. prepared the samples and developed the staining protocols. E.S. and H.C. contributed to optical modeling and system implementation. C.G. advised computational methodology. E.S., C.G., and H.C. supervised the project. All authors analyzed the data, discussed the results, and contributed to the manuscript.

Data availability

The MDCK fluorescence dataset acquired in this work has been deposited in Zenodo: <https://zenodo.org/records/18609221>. Third-party data from the Hyperspectral Image Dataset are publicly available as cited in the manuscript. Additional data supporting the findings of this study are available from the corresponding authors upon reasonable request.

Code availability

Training and inference code for the reconstruction network, together with pretrained model weights and example configuration files, are available on GitHub: <https://github.com/humeyracaglayan/MANTIS>. A black box version of the optical forward model, which maps ground-truth images to simulated sensor measurements, is also provided. The differentiable implementation of the optics forward model is available from the authors upon reasonable request, subject to intellectual property constraints.

Conflict of interest

The authors declare no competing interests.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41377-026-02337-y>.

Received: 4 December 2025 Revised: 12 April 2026 Accepted: 28 April 2026
Published online: 19 May 2026

References

- Liu, S. & Hua, H. Extended depth-of-field microscopic imaging with a variable focus microscope objective. *Opt. Express* **19**, 353–362 (2011).
- Glaser, A. K. et al. Light-sheet microscopy for slide-free non-destructive pathology of large clinical specimens. *Nat. Biomed. Eng.* **1**, 0084 (2017).
- Power, R. M. & Huiskens, J. A guide to light-sheet fluorescence microscopy for multiscale imaging. *Nat. Methods* **14**, 360–373 (2017).
- Böhm, U. L. & Judkewitz, B. Fast and light-efficient remote focusing for volumetric voltage imaging. *Nat. Commun.* **15**, 9555 (2024).
- Prince, N. H., Sain, N. & Chakraborty, T. Remote refocusing for multi-scale imaging. *J. Biomed. Opt.* **29**, 080501 (2024).
- Botcherby, E. J. et al. Aberration-free optical refocusing in high numerical aperture microscopy. *Opt. Lett.* **32**, 2007–2009 (2007).
- Broxton, M. et al. Wave optics theory and 3-D deconvolution for the light field microscope. *Opt. Express* **21**, 25418–25439 (2013).
- Gregor, I. et al. Instant three-color multiplane fluorescence microscopy. *Bio-phys. Rep.* **1**, 100001 (2021).
- Wang, P. et al. A single-shot hyperspectral phasor camera for fast, multi-color fluorescence microscopy. *Cell Rep. Methods* **3**, 100441 (2023).
- Zhang, C. S. et al. Deep tissue super-resolution imaging with adaptive optical two-photon multifocal structured illumination microscopy. *Photonix* **4**, 38 (2023).
- Icha, J. et al. Phototoxicity in live fluorescence microscopy, and how to avoid it. *BioEssays* **39**, 1700003 (2017).
- Dowski Jr, E. R. & Cathey, W. T. Extended depth of field through wave-front coding. *Appl. Opt.* **34**, 1859–1866 (1995).
- Cathey, W. T. & Dowski, E. R. New paradigm for imaging systems. *Appl. Opt.* **41**, 6080–6092 (2002).
- Akpinar, U. et al. Learning wavefront coding for extended depth of field imaging. *IEEE Trans. Image Process.* **30**, 3307–3320 (2021).
- Sitzmann, V. et al. End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging. *ACM Trans. Graph.* **37**, 114 (2018).
- Wu, Y. C. et al. PhaseCam3D — learning phase masks for passive single view depth estimation. In *Proc. 2019 IEEE International Conference on Computational Photography (ICCP)* 1–12 (IEEE, 2019).
- Jin, L. B. et al. Deep learning extended depth-of-field microscope for fast and slide-free histology. *Proc. Natl. Acad. Sci. USA* **117**, 33051–33060 (2020).
- Sahin, E. et al. Learning extended depth of field hyperspectral imaging. In *Proc. 2023 IEEE International Conference on Image Processing (ICIP)* 1850–1854 (IEEE, 2023).
- Yu, N. F. & Capasso, F. Flat optics with designer metasurfaces. *Nat. Mater.* **13**, 139–150 (2014).
- Khorasaninejad, M. et al. Metalenses at visible wavelengths: diffraction-limited focusing and subwavelength resolution imaging. *Science* **352**, 1190–1194 (2016).
- Huang, L. C. et al. Design and analysis of extended depth of focus metalenses for achromatic computational imaging. *Photonics Res.* **8**, 1613–1623 (2020).
- Whitehead, J. E. M. et al. Fast extended depth of focus meta-optics for varifocal functionality. *Photonics Res.* **10**, 828–833 (2022).
- Xiao, S. Y. et al. Inverse design of a near-infrared metalens with an extended depth of focus based on double-process genetic algorithm optimization. *Opt. Express* **31**, 8668–8681 (2023).
- Colburn, S., Zhan, A. L. & Majumdar, A. Metasurface optics for full-color computational imaging. *Sci. Adv.* **4**, 2114 (2018).
- Bayati, E. et al. Inverse designed extended depth of focus meta-optics for broadband imaging in the visible. *Nanophotonics* **11**, 2531–2540 (2022).
- Lin, R. J. et al. Achromatic metalens array for full-colour light-field imaging. *Nat. Nanotechnol.* **14**, 227–231 (2019).
- Fan, Z. B. et al. A broadband achromatic metalens array for integral imaging in the visible. *Light Sci. Appl.* **8**, 67 (2019).



28. Xu, B. B. et al. Metalens-integrated compact imaging devices for wide-field microscopy. *Adv. Photonics* **2**, 066004 (2020).
29. Seong, B. et al. E2E-BPF microscope: extended depth-of-field microscopy using learning-based implementation of binary phase filter and image deconvolution. *Light Sci. Appl.* **12**, 269 (2023).
30. Jin, L. B. et al. DeepDOF-SE: affordable deep-learning microscopy platform for slide-free histology. *Nat. Commun.* **15**, 2935 (2024).
31. Zhang, Y. L. et al. Large depth-of-field ultra-compact microscope by progressive optimization and deep learning. *Nat. Commun.* **14**, 4118 (2023).
32. Atalay Appak, I. A. et al. Learning flat optics for extended depth of field microscopy imaging. *Nanophotonics* **12**, 3623–3632 (2023).
33. Ronneberger, O., Fischer, P. & Brox, T. U-Net: convolutional networks for biomedical image segmentation. In *Proc. 18th International Conference on Medical Image Computing and Computer-Assisted Intervention* 234–241 (Springer, 2015).
34. Imamoglu, N. et al. Hyperspectral image dataset for benchmarking on salient object detection. In *Proc. 2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*. 1–3 (IEEE, 2018).
35. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. In 3rd International Conference on Learning Representations, Conference Track Proceedings (ICLR, 2015).
36. Tse, J. R. & Engler, A. J. Preparation of hydrogel substrates with tunable mechanical properties. *Curr. Protoc. Cell Biol.* **47**, 10.16.1–10.16.16 (2010).

