Research Article

Ipek Anil Atalay Appak, Erdem Sahin, Christine Guillemot and Humeyra Caglayan*

Learning flat optics for extended depth of field microscopy imaging

https://doi.org/10.1515/nanoph-2023-0321
Received May 28, 2023; accepted July 23, 2023;
published online August 2, 2023

Abstract: Conventional microscopy systems have limited depth of field, which often necessitates depth scanning techniques hindered by light scattering. Various techniques have been developed to address this challenge, but they have limited extended depth of field (EDOF) capabilities. To overcome this challenge, this study proposes an end-to-end optimization framework for building a computational EDOF microscope that combines a 4f microscopy optical setup incorporating learned optics at the Fourier plane and a post-processing deblurring neural network. Utilizing the end-to-end differentiable model, we present a systematic design methodology for computational EDOF microscopy based on the specific visualization requirements of the sample under examination. In particular, we demonstrate that the metasurface optics provides key advantages for extreme EDOF imaging conditions, where the extended DOF range is well beyond what is demonstrated in state of the art, achieving superior EDOF performance.

Keywords: diffractive optics; end-to-end learning; extended depth of field; metasurfaces; microscopy imaging

1 Introduction

A conventional imaging system can produce sharp images for objects within the depth of field (DOF), which is the range around the focused depth of the scene. The DOF coverage is inversely proportional to the numerical aperture (NA), i.e., a smaller NA leads to a larger DOF. Although specific applications may require smaller DOFs, a larger DOF is often preferable to obtain sharp images of objects at varying depths. In particular, microscopy requires high NA objectives to capture precise details surrounding the focused depth of an object. However, due to the shallow DOF of high NA imaging systems, microscopic imaging systems often use depth scanning techniques to cover the entire depth range of interest, typically much larger than the imaging system DOF [1]. Depth scanning techniques are often insufficient due to the light scattering from objects outside the intended image plane, resulting in artifacts or severe noise. Acquiring cross-sectional data using this approach requires sweeping a focused point across the entire sample, which inherently imposes temporal limitations on the frame rate and prevents snapshot acquisition. To address these challenges, various techniques have been developed, such as decoupled illumination and detection in light-sheet microscopy [2], dynamic remote focusing [3, 4], and spatial and spectral multiplexing [5, 6]. Despite their potential advantages, these methods often necessitate a specialized and intricate optical configuration, which may render them both costly and difficult to implement in microscopy applications. In addition, computational approaches such as Fourier ptychographic microscopy demonstrated extended depth of field (EDOF) imaging capabilities [7]. However, this method relies on image reconstruction that assumes a thin sample illuminated by oblique plane waves, rendering it unsuitable for clinical fluorescence imaging applications.

The integration of wavefront encoding with computational reconstruction methods offers a cost-effective and efficient approach to improve EDOF imaging performance. Specifically, computational reconstruction methods, such as

***Corresponding author: Humeyra Caglayan**, Faculty of Engineering and Natural Science, Photonics, Tampere University, 33720 Tampere, Finland, E-mail: humeyra.caglayan@tuni.fi. https://orcid.org/0000-0002-0656-614X
Ipek Anil Atalay Appak, Faculty of Engineering and Natural Science, Photonics, Tampere University, 33720 Tampere, Finland; and INRIA Rennes – Bretagne Atlantique, Rennes, France, E-mail: anil.atalay@tuni.fi. https://orcid.org/0000-0001-5327-7423
Erdem Sahin, Faculty of Information Technology and Communication Sciences, Tampere University, 33720 Tampere, Finland, E-mail: erdem.sahin@tuni.fi. https://orcid.org/0000-0002-5371-6649
Christine Guillemot, INRIA Rennes – Bretagne Atlantique, Rennes, France, E-mail: christine.guillemot@inria.fr. https://orcid.org/0000-0003-1604-967X

deep learning-based approaches and iterative optimization algorithms, have demonstrated substantial advancements in enhancing the EDOF imaging performance by effectively mitigating defocus blur, noise, and artifacts [8–12]. These methods leverage the optimization of both optical components and the reconstruction algorithm, enabling the system to enhance the EDOF performance effectively. End-to-end optimization frameworks, which jointly optimize the optical design and the associated reconstruction algorithms, have emerged as a powerful tool for addressing the EDOF challenge. This approach allows the system to learn and adapt to specific imaging requirements, thereby improving overall performance and enabling better control over trade-offs in resolution, noise, and depth range [8]. Wavefront coding involves the use of a phase element, such as a diffractive optical element (DOE) or a free-form refractive lens, placed at the aperture plane [13–19]. The main objective of EDOF wavefront coding is to achieve a depth-invariant point spread function (PSF) while preserving information at all spatial frequencies. The research presented in [10] employs DOEs to realize EDOF microscopy across 200 μm DOF range, a methodology closely aligned with our study. However, DOE wavefront coding displays restricted EDOF capabilities, a limitation that may be attributed to constraints within the space-bandwidth product (SBP). This factor determines the information content captured by the imaging system. A detailed comparison with such method, called DeepDOF, is elaborated in Section 3.

In comparison, metasurfaces, which are ultra-thin meta-optics composed of subwavelength nano-antennas, offer increased design flexibility and a superior SBP compared to DOEs [20, 21]. These structures facilitate precise control over phase, amplitude, and polarization of light at the nanoscale, allowing almost arbitrary modification of the complex optical functions on a thin, planar device. The advantages of metasurfaces can be attributed to the rich modal characteristics of meta-optical scatterers, enabling multifunctional capabilities beyond traditional DOEs, encompassing polarization, frequency, and angle multiplexing [22–25]. Consequently, metasurfaces exhibit a greater potential for addressing the EDOF microscopy challenge more effectively than conventional DOEs, with researchers having already leveraged their benefits in various applications such as flat optics for imaging [26, 27], polarization control [28], and holography [29]. Despite the promising potential of meta-optics, current metasurface imaging methodologies demonstrate limited EDOF imaging capabilities. The EDOF range for any given imaging system can be quantified using the defocus coefficient, wherein a larger EDOF corresponds to a higher defocus coefficient. Details regarding these defocus coefficients and their relationship with the depth of field range and optical parameters are comprehensively covered in Section 2.1. Currently, most methods are designed to accommodate systems with a maximum defocus coefficient limited to around 75, with some using mechanical displacement as a strategy to extend imaging capability [30–32]. Although the narrow defocus ranges may be adequate for certain applications, it is comparatively limited in broader scientific and industrial contexts. As such, there remains a need for more flexible and versatile imaging solutions that can accommodate a broader depth range without sacrificing image quality. In contrast, the study in [33], addresses a problem characterized by a maximum defocus coefficient comparable to ours, valued at around 245. However, the demonstrated image quality is considerably lower compared to the results achieved in our research. It is worth noting that other metasurface-based imaging implementations exist, which are capable of encoding spatial, spectral, and polarization information while maintaining satisfactory imaging performance [34, 35]. Moreover, computational metasurface designs that yield a large field-of-view for full-color metasurface operation, without significant degradation of imaging performance, have been reported [21].

In this study, we propose an end-to-end optimization framework designed for acquiring high-resolution images across an extensive DOF range within a microscopy system. The optics and post-processing algorithm are modeled as parts of the end-to-end differentiable computational image acquisition system, allowing for simultaneously optimizing both components. Our computational EDOF microscope employs a hybrid approach that combines a 4f microscopy optical setup with a learned wavefront modulating optical element at the Fourier plane. We explore metasurfaces and (conventional) DOEs for implementing such novel design modulation. The encoded image acquired at the sensor is post-processed by a convolutional neural network (CNN) that implements deblurring to achieve an EDOF image of the specimen. The optimization procedure involves tuning of critical parameters within two primary components of the system: the optical component, characterized by the phase modulation function, and the image reconstruction component, realized through the deblurring convolutional neural network (D-CNN). This process is facilitated by an end-to-end learning methodology that utilizes a dataset of sharp images to steer the optimization. This methodology emphasizes the importance of carefully outlining sampling requirements for various depth-of-field targets in order to achieve optimal imaging results. As a result, our systematic design methodology significantly outperforms existing state-of-the-art EDOF imaging techniques in terms of image quality and depth range.

2 Methods

The EDOF microscopy problem addresses the difficulty of capturing high-resolution images across an extensive DOF range within a microscopy system. The objective of resolving this problem is to improve the imaging performance of the system by expanding its DOF, allowing the acquisition of sharp images over a larger depth range without necessitating mechanical focus adjustments. Figure 1 presents the proposed 4f system designed to address the EDOF microscopy problem, which consists of an optical module, a sensor, and a subsequent deblurring module for post-processing. The optical module employs a 4f imaging configuration that incorporates a phase coding mask at the aperture position. This mask, an optical element crafted to modulate incident light, introduces phase shifts across various regions through a transparent material with spatially varying geometry. The purpose of using a phase coding mask is to manipulate the optical wavefront, enabling a specific PSF or other desired properties within the optical system. In addressing this problem, the 4f system components remain fixed while optimizing the EDOF microscopy imaging system by learning the spatial distribution of the phase coding mask and the D-CNN weights for the targeted DOF ranges. Consequently, the objective is to achieve a defocus-invariant PSF within the optical system, resulting in improved EDOF microscopy imaging performance.

Figure 2 depicts the proposed end-to-end learning procedure for the EDOF microscopy problem. This process accepts two inputs as high-resolution image patches and the predetermined depth range. The architecture of the framework consists of two main components: an optical layer and a D-CNN layer. During the training phase, the sensor image is generated through the optical layer, which involves simulating the image acquisition process based on depth value and the input image. The resulting sensor image is then fed into the D-CNN, which estimates the sharp, deblurred image as its output. The backpropagation algorithm is used to update the spatial phase distribution of the phase coding mask and the weights of the D-CNN at each iteration of the learning procedure, thus optimizing the parameters of the EDOF microscope in an end-to-end manner. The optimized phase coding mask is realized as a meta-optic element (metalens) or DOE, depending on the spatial phase distribution and sampling. After training is completed for the targeted DOF range, the defocus-invariant optical element and D-CNN model are obtained in the 4f system computational EDOF microscope model (Figure 2). Further details regarding the optical and D-CNN layers are provided below.

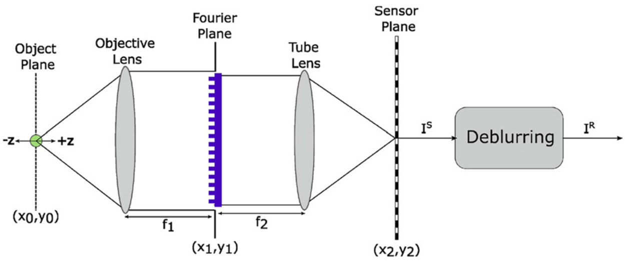

Figure 1: A 4f system computational EDOF microscope model that combines the optical module, sensor, and D-CNN.

2.1 Optical layer

The optical layer is a computational module that operates based on the principles of wave optics and performs sensor image formation. In the following, we discuss sensor image formation by employing wave optics in particular and characterize the parameters of the EDOF microscope for the targeted DOF range for a 4f shift-invariant imaging system as illustrated in Figure 1. Here, the specimen is illuminated by a monochromatic, spatially incoherent light source. The spatial coordinates in the Fourier and sensor planes are denoted as (x_1, y_1) and (x_2, y_2), respectively. A transparent biological specimen in the system can be represented as a stack of 2D images corresponding to a fixed scene depth where a slice of such a stack corresponds to $I_z(x_2, y_2)$ at the scene depth z. The contribution of such a slice to the sensor image, $I_z^s(x_2, y_2)$, is determined by convolution with the depth-dependent PSF, $h_z(x_2, y_2)$:

$$I_z^s(x_2, y_2) = I_z(x_2, y_2) * h_z(x_2, y_2). \tag{1}$$

The final image captured at the sensor, $I^s(x_2, y_2)$, is then determined by integrating $I_z^s(x_2, y_2)$ over all depth values possible in the scene, considering

$$I^s(x_2, y_2) = \int I_z^s(x_2, y_2)\, \mathrm{d}z + \eta_s, \tag{2}$$

where η_s is a sensor noise. In our simulations, we consider the noise as a zero-mean Gaussian model with $\eta_s \approx N\left(0, \sigma_s^2\right)$, where σ_s represents the standard deviation of the Gaussian noise. It is important to note that while we have assumed Gaussian noise for our experiments, the proposed method can be easily adapted to handle other noise models, such as Poisson distributed noise, by adjusting the noise assumption within the optimization process. Considering Eqs. 1 and 2, the recovery of a sharp image directly depends on $h_z(x_2, y_2)$. The PSF on the sensor plane can be modeled using Fourier optics as the square of the Fourier transform of the generalized pupil function:

$$h_z(x_2, y_2) = |F\{P(x_1, y_1)\}|^2, \tag{3}$$

where $F\{.\}$ denotes the Fourier transform operator, and $P(x_1, y_1)$ is the pupil function, which describes the relative amplitude and phase changes of the wavefront at the Fourier plane:

$$P(x_1, y_1) = A(x_1, y_1)\mathrm{e}^{\mathrm{i}\phi(x_1, y_1)}. \tag{4}$$

The pupil can be modulated using a phase coding mask introducing a phase term $\phi^M(x_1, y_1)$ to enhance the defocus invariance of the PSF in the targeted DOF range. The resultant phase term becomes:

$$\phi_z(x_1, y_1) = \phi_z^{DF}(x_1, y_1) + \phi^M(x_1, y_1). \tag{5}$$

The defocus aberration due to the mismatch between in-focus depth z_0 and the actual depth z of a scene point is

$$\phi_z^{DF}(x_1, y_1) = \psi_z \frac{x_1^2 + y_1^2}{r^2}, \tag{6}$$

where $\psi_z = \frac{\pi}{\lambda}\left(\frac{1}{z} - \frac{1}{z_0}\right)r^2$ is the defocus coefficient and r is the radius of the pupil. To represent all spatial frequencies k_x, k_y supported by the objective lens on the pupil plane, the pupil size must be chosen using

$$r \geq \frac{f_1}{k}\sqrt{k_x^2 + k_y^2}, \tag{7}$$

where f_1 is the focal length of the tube lens, and k is the wave number. The frequency support of the PSF, in particular, determines the reconstruction quality. To avoid aliasing due to undersampling of the defocused pupil function, which would otherwise result in a miscalculated

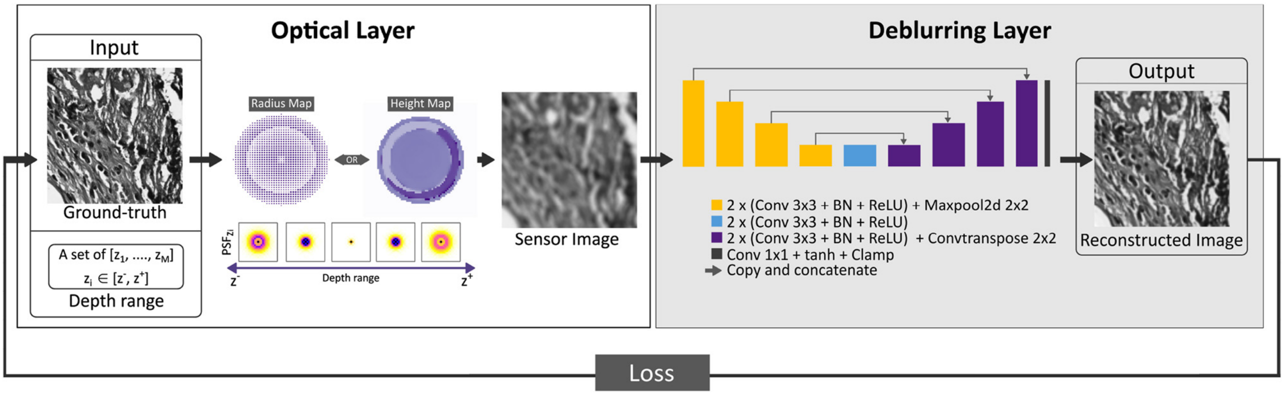

Figure 2: Overall representation of the end-to-end learning framework for joint optimization of the phase coding mask and the D-CNN. The end-to-end framework consists of an optical layer that simulates image formation for the learned optics, and the deblurring layer employs U-Net architecture to reconstruct in-focus images within the desired DOF range. The optical element can be realized via a metalens, parametrized by the radius map, or a DOE, parametrized by the height map. The choice of representation relies upon the maximum Δs calculated for the intended depth range.

PSF, the minimum sampling rate of the simulated phase coding mask must be determined. The sampling rate Δs for a given depth range should satisfy [11]

$$\Delta s \leq \frac{r\pi}{8\psi_{\max}}, \tag{8}$$

where $\psi_{\max} = \max\left(|\psi_{z^-}|, |\psi_{z^+}|\right)$, the maximum defocus value within the scene. At the same time, we use such a relation as a guide in choosing the search space for the phase modulation function of the optical element, matching its degree of freedom with the complexity of the EDOF imaging problem at hand. This is observed to have a significant help in the end-to-end learning process, i.e., its convergence to the optimal solution. The optimized optical element can be realized as a metalens or DOE according to the calculated maximum Δs for the targeted depth range. The metalens and DOE are parameterized by a radius map and a height map, respectively, details of which are provided in the following sections.

During training, the optical element is optimized to minimize the impact of the defocus phase and ensure depth-invariant PSFs through the utilization of learned phase coding mask elements. For each iteration within the forward pass, this optimization is achieved by calculating Eq. (3) and optimizing the $\phi^M(x_1, y_1)$ phase modulation term. For both the DOE and the metalens, the amplitude $A(x_1, y_1)$ of the pupil function is kept constant within the aperture diameter and is modeled as a circular function. The camera parameters f_1, r, and Δs are predefined as optical parameters. r and Δs is calculated for the selected objective lens and targeted depth range using Eqs. (7) and (8), respectively.

For the DOE design, the phase modulation can be controlled through unit cell height. Specifically, the phase shift is given by the equation:

$$\phi^M(x_1, y_1) = k(n-1)h(x_1, y_1), \tag{9}$$

where n denotes the refractive index of the DOE material which is 1.5, specifically at the design wavelength of 550 nm within the green spectrum. The DOE is assumed to be lossless; therefore the amplitude is kept fixed as the circular function within the aperture diameter.

For the metalens design, phase accumulation is achieved through the waveguiding effect [36], whereby the height of the nanopillars is selected to provide 2π phase coverage across a range of radii. While the smallest possible diameter is primarily limited by fabrication constraints, the largest diameter is 50 nm smaller than Δs, as set by Eq. (8). To ensure high efficiency, the nanopillar height Δh is optimized at the 550 nm design wavelength. The phase and transmission responses are simulated via a finite-difference time-domain (FDTD) analysis, which involves varying the radius of the gallium nitride (GaN) nanopillar on a sapphire glass substrate. As depicted in Figure 3(a), the analysis results in full phase coverage (0–2π) with high transmission (overall greater than 83 %) for 550 nm.

Irrespective of the desired target phase, designing a metalens involves converting a spatial phase profile into a corresponding spatial radius distribution. A resulting $\phi^M(x_1, y_1)$ phase profile, generated from the end-to-end framework, is illustrated in Figure 3(b). This profile is then translated into the full metalens, taking into account the simulated phase and radius response. This approach enables the design of highly precise and effective metalens that meets the desired phase requirements.

2.2 Deblurring CNN

The D-CNN module shown in Figure 2 utilizes the sensor output (I^s) from the optical layer as its input. Although many network architectures exist for this problem, we chose the well-known U-Net [37] as it is widely used in biomedical imaging for image reconstruction [10, 38, 39]. In short, the U-Net implementation has an encoder and decoder architecture with 23 convolution layers and 32 to 512 feature channels. At each step of the encoder stage, the input underwent two 3×3 convolution layers, a rectified linear unit (ReLU), and batch normalization (BN). Subsequently, the feature map is downscaled by a 2×2 max-pooling operation. Likewise, following two 3×3 convolution layers that incorporated ReLU and BN at each iteration of the decoder, a 2×2 transposed convolution operator upsampled the feature map. At the final layer, a 1×1 convolution is used to map each 32-component feature vector to the desired number of classes, which is 1 in the current case. Moreover, a hyperbolic tangent (tanh) activation is utilized to map the output to the range of $[-1, 1]$. The residual image is then incorporated by addition to the sensor output. The Clamp layer further processes the resulting image to constrain the data from 0 to 1.

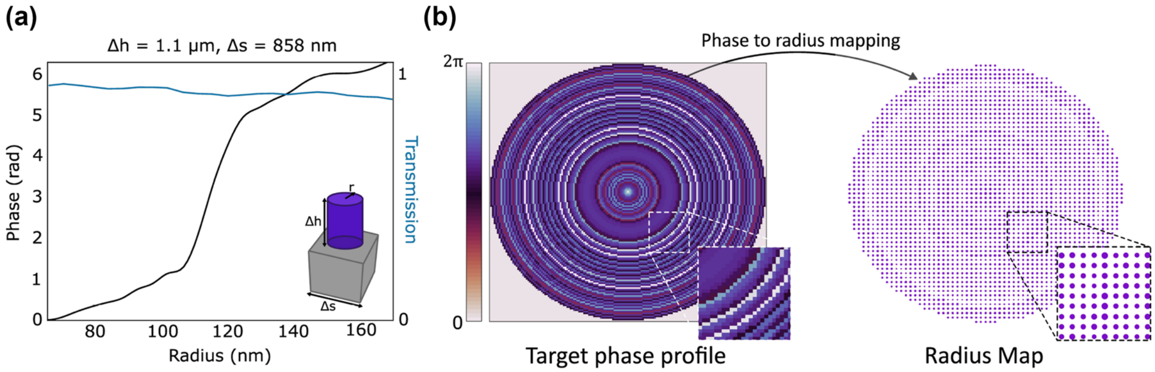

Figure 3: Design and simulation of EDOF imaging metalenses. The metalenses are made up of GaN nanopillars with a refractive index of 2.42, where the thickness (Δh), sampling (Δs), and radius (Δr) are the design parameters. (a) Simulation of the nanopillars' transmission amplitude and phase response via FDTD. (b) Illustration of the full metalens design, transforming the spatial phase profile into a spatial radius (of nanopillar) distribution with respect to the target phase.

For memory efficiency, U-Net takes input as 256 × 256 blurry pixel images corresponding to different depths and outputs the enhanced images at each iteration. It should be noted that to avoid border distortions during convolution with the PSF, the patched image and PSF are convolved and subsequently cropped to 256 × 256 after convolution. Once the network is trained, it can process images with dimensions that are multiples of 16 in width and height.

2.3 End-to-end learning

The end-to-end framework is trained via an image dataset provided by [10]. The dataset consists of a diverse array of microscopic fluorescent images of proflavin-stained oral cancer resections, histopathology images of healthy and cancerous tissues from the Cancer Genome Atlas (TCGA), and natural images from the National Digital Science and Technology Research Institute's (INRIA) Holiday dataset. This diverse selection ensures a broad range of feature scales within the dataset. The dataset contains a total of 1800 grayscale images, 1000 × 1000 pixels each, and includes 600 images of each type. The images are randomly assigned to training, validation, and testing datasets, with 1500, 150, and 150 images, respectively. During training, the images are randomly cropped and enhanced with rotation, flipping, and brightness adjustments for data augmentation. Throughout the training and validation stages, at each iteration, a random image patch is selected from the corresponding dataset and assigned to M depth positions $z_i \in (z_1, \dots, z_M)$ that are uniformly distributed (in diopters) within the boundaries of the targeted scene depth, as illustrated in Figure 2. During a forward pass, the loss function corresponding to each depth is computed and subsequently summed and averaged over the number of selected depths. The calculation of the loss function for each depth involves sequentially processing the image patches assigned to the chosen z_i. In this study, we limited m to five depth positions to reduce the computational complexity. Upon experimentation, an increase in the number of depth positions did not yield a significant improvement in imaging performance.

The loss function used in the framework is the root mean squared error (RMSE) calculated for each depth and each pixel of the reconstructed image stack, compared to the respective pixel in the ground-truth image, with the results subsequently averaged across the number of depths:

$$L_{\text{RMSE}} = \frac{1}{M}\sum_{i=1}^{M}\frac{1}{\sqrt{N}}\|I - I^R\|_2, \tag{10}$$

where N is the number of pixels. As the input experiences blurring at varying defocus values throughout the depth, the framework intrinsically adapted the optics to achieve a defocus-insensitive PSF. Consequently, no explicit cost function is required to maintain PSF similarity within the designated depth range.

The PyTorch package is used to implement the framework, and stochastic gradient descent with the Adam optimizer [40] is employed for optimization. The learning rates for the Adam optimizer are selected empirically as $1e-7$ for the optical layer and $1e-4$ for the D-CNN. During end-to-end framework training, a two-step training process is used, where the initial step is training the U-Net with fixed optics. After the convergence, joint training of the optical rendering and D-CNN is performed to achieve optimal performance. Regarding computational requirements, we used two Tesla P100-PCIE-12GB GPUs for training. Depending on the complexity of the problem, which is dictated by the number of required parameters, the training duration varies. The most parameter-intensive problem necessitated a training period of approximately 148 h and 33 min, whereas the least demanding scenario required a comparatively shorter duration of around 6 h and 41 min.

3 Results

Our design is aimed to mitigate the trade-off between DOF and spatial resolution for varying target DOFs by experimenting with different objective lenses within the 4f imaging setup, as illustrated in Figure 1. This trade-off is mathematically expressed as follows:

$$\text{DOF} \propto \frac{\lambda}{\text{NA}^2} \propto \frac{\text{resolution}^2}{\lambda}. \tag{11}$$

This demonstrates that the spatial resolution is higher for a shallower DOF range; therefore, a shallower DOF range optimization problem is comparatively less challenging to solve. The complexity of the problem, however, is affiliated with the selected lenses and the desired DOF range. It is quantified and expressed using the defocus coefficient (ψ_z) and presented alongside the system parameters for different simulations in Table 1.

Table 1 demonstrates that solving the EDOF problem for a higher NA lens system is considerably more challenging task, where the sampling requirements, determined by Eq. (8), become more demanding due to the broader bandwidth of the defocused pupil function, necessitating finer sampling for accurate results. Such a case, thus, advocates the metalens to address the underlying stricter sampling requirement. Conversely, for the same EDOF range, using a low NA lens-based system with a learned DOE can yield a satisfactory solution. However, it should be noted that the final magnification would be significantly lower in such a system. Therefore, the lens selection should be based on the requirements of the sample to be visualized. To the best of our knowledge, certain biological specimens, such as the developing embryo, exhibit a diameter that increases from approximately 200 μm to nearly 3 mm [41]. As such, we have selected to target DOF ranges of 200 μm, 2 mm, and 3 mm for our study.

To further support the results, we compare the proposed algorithm with two existing methods. Initially, a cubic mask is adopted as the conventional method for DOF extension [13], which modulates the phase as

$$\phi^M(x_1, y_1) = \mathrm{mod}\left[\frac{\alpha}{N^3}\left(x_1^3 + y_1^3\right), 2\pi\right], |x| < \frac{N}{2}, |y| < \frac{N}{2}, \tag{12}$$

where α determines the number of 2π transitions. In this case, a fixed cubic phase mask is combined with a U-Net and subsequently trained to perform deblurring for all EDOF targets. To select an appropriate mask for each DOF range, the modulation transfer function (MTF) is evaluated across five different depths by tuning α, to be insensitive to defocus and represent all spatial frequencies. α values of 6π, 200π, and 360π are selected for the DOF ranges of 200 μm, 2 mm, and 3 mm, respectively.

Additionally, the DeepDOF algorithm proposed by Jin et al. [10] is adopted as a more recent advanced EDOF method based on the end-to-end learning framework and Zernike basis (Z_n) representation. Notably, the paper does not provide a relationship between the choice of the number of Zernike polynomials and the EDOF range. To align the optical setup parameters with our methodology and ascertain the appropriate number of parameters (n) for each depth range, we calculated the Fourier transform of the Z_n and found the bandwidth of the resultant signal. Considering the Nyquist theorem, we calculated the sampling rate for the Z_n. Based on these calculations, we selected the first 11, 1100, and 2000 Z_n for DOF targets of 200 μm, 2 mm, and 3 mm, respectively.

We present a quantitative analysis of the proposed method in comparison with existing approaches by analyzing the performance metrics, specifically peak signal-to-noise ratios (PSNRs) and structural similarities (SSIMs). For each test setup, the PSNR and SSIM values are derived and depicted in Figure 4. Additionally, the mean PSNR and SSIM values are calculated for all images within the 150 test images of the dataset and presented in Table 2. To calculate the average values, we assign each test image to predetermined depths, uniformly distributed throughout the scene depth range. Subsequently, the PSNR and SSIM values are averaged.

As demonstrated in Table 2, the proposed method outperforms existing approaches in terms of PSNR and SSIM values, both for the sample image shown in Figure 4 and the mean values across the entire test dataset. In particular, the D-CNN alone, as inferred from the fixed Cubic mask and U-Net simulations, is inadequate. Upon increasing the target DOF, the image becomes excessively blurry at the sensor level, and the U-Net cannot recover the images effectively. Comparable results are visible in the DeepDOF approach, which is linked to the sub-sampling of the frequency domain that arises from using the Zernike basis representation. It is important to acknowledge that [10] is optimized using 55 Zernike basis, for a target DOF of 200 μm. The starred result in Table 2 displays the outcomes obtained when employing 55 Zernike basis. In this case, we retain the same setup parameters as in our other simulations, only modifying the number of Zernike basis utilized to represent the height map in the method, resulting in a higher sampling than necessary. The results demonstrate that increasing the sampling does not improve the performance, thus confirming that our calculated setup parameters sufficiently address

Table 1: Test results and calculated system parameters for the selected lenses.

Objective lens	NA	DOF	r (mm)	Δs(μm)	PSNR	SSIM	$\psi_{\max}$
RMS4X-PF[a]	0.13	±100 μm	3.72	740	38.19	0.98	0.97
RMS4X-PF	0.13	±1 mm	3.72	73	30.11	0.88	9.877
RMS20X-PF	0.40	±100 μm	2.87	38	29.95	0.90	14.69
Mitutoyo50X	0.42	±100 μm	1.07	20	30.84	0.88	10.48
Mitutoyo50X[a]	0.42	±1 mm	1.07	1.54	29.28	0.89	136.24
Mitutoyo50X[a]	0.42	±1.5 mm	1.07	0.858	27.98	0.89	245.54

[a] Used as the optimized design for the targeted DOF range.

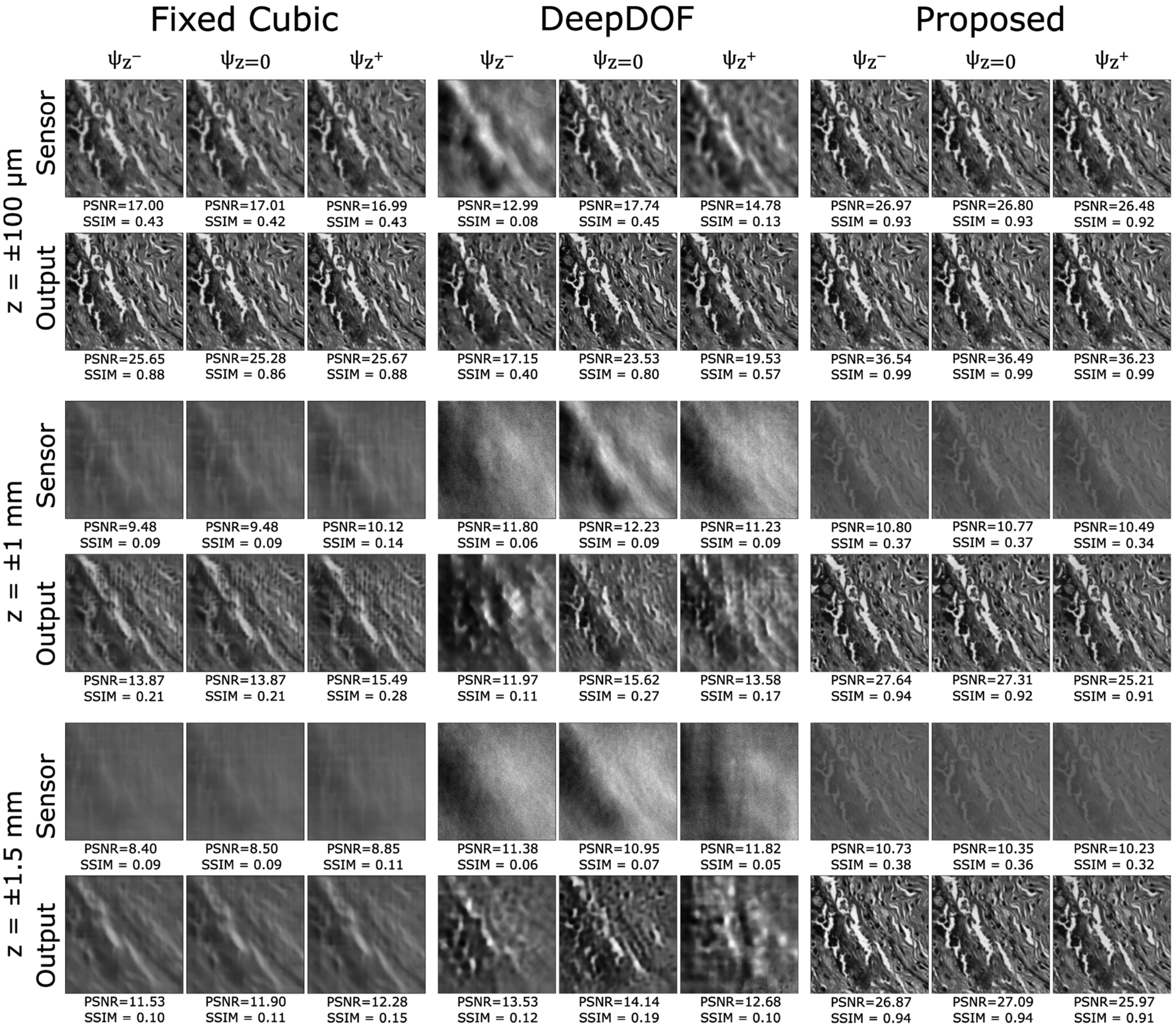

Figure 4: Simulated performance of a cubic-mask-enabled computational 4f system, DeepDOF model, computational EDOF microscope, optimized for various target depths. For all DOFs, our method shows superior performance.

the system complexity. Additionally, in our U-Net implementation, different from [10], we adapted a modified U-Net architecture for the deblurring layer. The main difference resides in the final layer of our U-Net architecture, where we incorporated the residual image and applied the clamp function, enhancing both the accuracy and efficiency of the deblurring process.

An alternative method for evaluating EDOF performance involves examining the characteristics of MTFs across various depths, with MTFs representing the magnitudes of the Fourier transforms corresponding to their associated PSFs [12]. To facilitate efficient depth-agnostic deblurring throughout the whole depth range of interest, MTF pass-bands should be as wide as possible to recover features at various spatial frequencies while maintaining a high degree of similarity among themselves. Figure 5 presents the final metasurface designs for the 3 mm DOF scenario, as well as the MTFs at three distinct depths. The depth dependency of the MTFs in the proposed method is decreased compared to both the DeepDOF and Fixed cubic cases. The MTFs of the proposed method display greater similarity among themselves and avoid crossing zero, ensuring the preservation of information during recovery. This similarity is attributed to depth-agnostic deblurring assisting MTFs to remain consistent across all targeted depths. To compare the frequency support of each method, we defined a threshold of 0.1 and examined the MTF magnitude across the frequency spectrum and depths, which formally

Table 2: Quantitative analysis of methods for EDOF problem. It should be noted that all the methods are simulated using the same system parameters that are computed for each DOF range. Refer to Table 1 for system parameters.

Framework configuration	DOF	PSNR	SSIM
Fixed cubic + U-Net	±100 µm	34.99	0.92
	±1 mm	23.6	0.62
	±1.5 mm	21.86	0.58
DeepDOF [10]	±100 µm	28.90, 28.85[a]	0.80, 0.81[a]
	±1 mm	20.67	0.55
	±1.5 mm	21.14	0.58
Our method	±100 µm	38.19	0.98
	±1 mm	29.28	0.89
	±1.5 mm	27.98	0.89

[a] Used first 55 Zernike polynomials.

determines the maximum spatial resolution transmitted through the optics. The MTFs of the proposed method resulted in broadband coverage across the entire spectrum, illustrating that even at higher spatial frequencies, MTFs do not approach zero. This finding corresponds to the clear visibility of small structures. Conversely, at lower spatial frequencies, the MTF is closer to 1, representing the ability to clearly visualize large structures. Indeed, as demonstrated in the other methods, all MTFs can clearly visualize large structures, but for smaller structures, quality is inadequate.

3.1 Fabrication error analysis

The robustness of the proposed EDOF microscope model is tested considering various levels of inaccuracies for the possible fabrication errors of metasurface and DOE. In particular, the effects of such fabrication inaccuracies are modeled by introducing random Gaussian-distributed noise with a standard deviation σ_r, to the optimized optical elements, as the radius and DOE height error during the test stage. The tested metasurface radius deviation levels are $\sigma_r =$ 5 nm, 12 nm, while the DOE height deviation is assumed as $\sigma_h = 30$ nm, 50 nm. Figure 6 presents the outcomes for the various EDOF configurations. As it can be inferred from the figure, the post-processing method performs adequately for lower fabrication error levels but exhibits limited robustness in the presence of higher error margins, leading to a noticeable reduction in image quality, particularly at the boundaries of the targeted EDOF. The robustness of the algorithm can be increased by incorporating fabrication error boundaries during training or retraining the D-CNN following the fabrication of optical components. Additionally, the implementation of advanced denoising algorithms can contribute to the enhancement of the results.

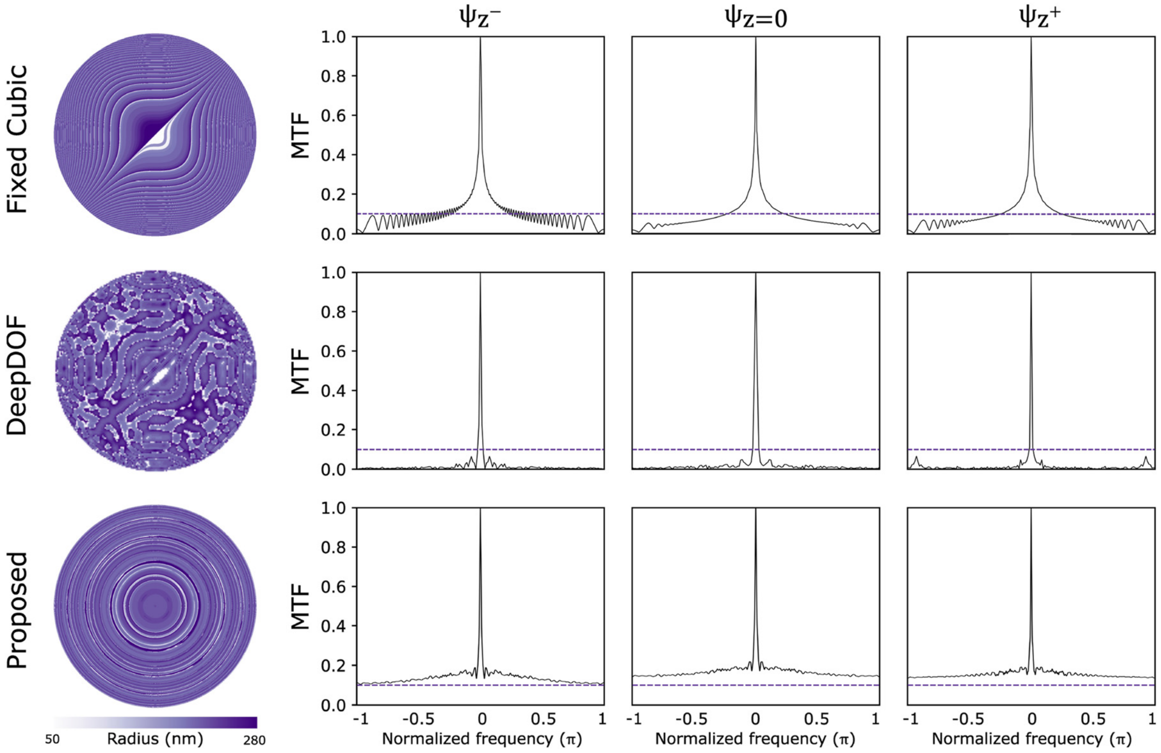

Figure 5: The optimized metalens radius maps, as well as the MTFs at three distinct depths, are presented for the existing and proposed methods, encompassing the maximum defocus values within the 3 mm DOF scene and the in-focus depth.

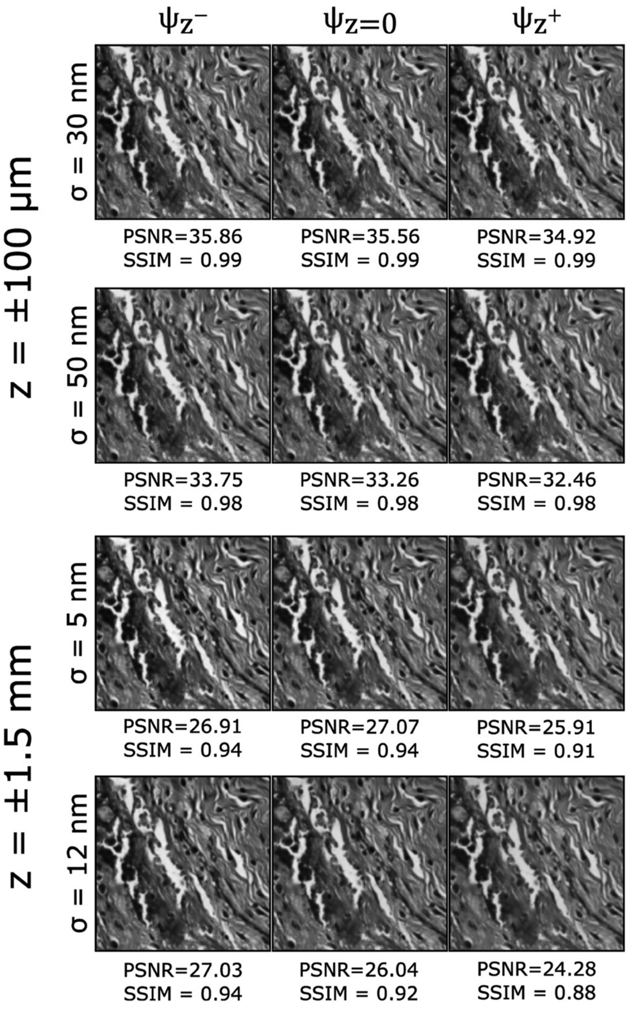

Figure 6: Reconstruction results for increasing radius and height-map fabrication error levels.

4 Future work and conclusion

In conclusion, we implemented an end-to-end optimization framework that effectively addresses the physical limitations inherent in the 4f system EDOF microscope. Our approach offers two primary advantages over existing methods. Firstly, the incorporation of metasurface optics within our system has facilitated the achievement of the most extensive EDOF range reported in the literature thus far. In extreme EDOF scenarios, specifically when employing a higher NA lens, meta-optics offer a distinct advantage due to their increased design flexibility and superior SBP. Secondly, our method effectively minimizes defocus by systematically incorporating relevant optical system parameters throughout the optimization process. The results obtained from our simulations demonstrate that our proposed technique outperforms the current state-of-the-art methods in EDOF microscopy imaging, delivering consistently high performance across a broad range of EDOF values.

In future work, we aim to investigate the co-design of optics and post-processing for broadband EDOF imaging. Moreover, we aim to explore the application of our approach in addressing the challenges associated with light field microscopy. Our next objective is fabricating a selected metasurface, followed by a comprehensive evaluation of its performance within the context of the 4f system setup. We believe that these research directions will contribute significantly to the ongoing advancement of microscopy imaging, ultimately leading to more sophisticated and versatile optical systems.

Research funding: This project has received funding from the European Union’s Horizon 2020 research and innovation programme 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).

Author contributions: All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Conflict of interest statement: The authors state no conflicts of interest.

Data availability: The datasets generated and/or analysed during the current study are available from the corresponding author upon reasonable request.

References

[1] S. Liu and H. Hua, “Extended depth-of-field microscopic imaging with a variable focus microscope objective,” *Opt. Express*, vol. 19, no. 1, pp. 353–362, 2011.

[2] A. K. Glaser, N. P. Reder, Y. Chen, et al., “Light-sheet microscopy for slide-free non-destructive pathology of large clinical specimens,” *Nat. Biomed. Eng.*, vol. 1, no. 7, p. 0084, 2017.

[3] W. J. Shain, N. A. Vickers, B. B. Goldberg, T. Bifano, and J. Mertz, “Extended depth-of-field microscopy with a high-speed deformable mirror,” *Opt. Lett.*, vol. 42, no. 5, pp. 995–998, 2017.

[4] S. Xiao, H. A. Tseng, H. Gritton, X. Han, and J. Mertz, “Video-rate volumetric neuronal imaging using 3D targeted illumination,” *Sci. Rep.*, vol. 8, no. 1, p. 7921, 2018.

[5] S. Abrahamsson, J. Chen, B. Hajj, et al., “Fast multicolor 3D imaging using aberration-corrected multifocus microscopy,” *Nat. Methods*, vol. 10, no. 1, pp. 60–63, 2013.

[6] S. Geissbuehler, A. Sharipov, A. Godinat, et al., “Live-cell multiplane three-dimensional super-resolution optical fluctuation imaging,” *Nat. Commun.*, vol. 5, no. 1, p. 5830, 2014.

[7] G. Zheng, R. Horstmeyer, and C. Yang, “Wide-field, high-resolution Fourier ptychographic microscopy,” *Nat. Photonics*, vol. 7, pp. 739–745, 2013.

[8] V. Sitzmann, S. Diamond, Y. Peng, et al., "End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging," *ACM Trans. Graph.*, vol. 37, pp. 1—13, 2018.

[9] Y. Wu, V. Boominathan, H. Chen, A. Sankaranarayanan, and A. Veeraraghavan, "PhaseCam3D — learning phase masks for passive single view depth estimation," in *2019 IEEE International Conference on Computational Photography* , ICCP, 2019, pp. 1—12.

[10] L. Jin, Y. Tang, Y. Wu, et al., "Deep learning extended depth-of-field microscope for fast and slide-free histology," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 117, pp. 33051—33060, 2020.

[11] U. Akpinar, E. Sahin, M. Meem, R. Menon, and A. Gotchev, "Learning wavefront coding for extended depth of field imaging," *IEEE Trans. Image Process.*, vol. 30, pp. 3307—3320, 2019.

[12] A. Levin, R. Fergus, F. Durand, and W. T. Freeman, "Image and depth from a conventional camera with a coded aperture," *ACM Trans. Graph.*, vol. 26, no. 3, pp. 70—es, 2007.

[13] E. R. Dowski and W. T. Cathey, "Extended depth of field through wave-front coding," *Appl. Opt.*, vol. 34, no. 11, pp. 1859—1866, 1995.

[14] V. N. Le, S. Chen, and Z. Fan, "Optimized asymmetrical tangent phase mask to obtain defocus invariant modulation transfer function in incoherent imaging systems," *Opt. Lett.*, vol. 39, no. 7, pp. 2171—2174, 2014.

[15] H. Zhao and Y. Li, "Optimized sinusoidal phase mask to extend the depth of field of an incoherent imaging system," *Opt. Lett.*, vol. 35, no. 2, pp. 267—269, 2010.

[16] S. S. Sherif, W. T. Cathey, and E. R. Dowski, "Phase plate to extend the depth of field of incoherent hybrid imaging systems," *Appl. Opt.*, vol. 43, no. 13, pp. 2709—2721, 2004.

[17] T. Stone and N. George, "Hybrid diffractive-refractive lenses and achromats," *Appl. Opt.*, vol. 27, no. 14, pp. 2960—2971, 1988.

[18] O. Cossairt and S. Nayar, "Spectral focal sweep: extended depth of field from chromatic aberrations," in *2010 IEEE International Conference on Computational Photography (ICCP)*, IEEE, 2010, pp. 1—8.

[19] H.-Y. Sung, S. S. Yang, and H. Chang, "Design of mobile phone lens with extended depth of field based on point-spread function focus invariance," in *Novel Optical Systems Design and Optimization XI*, vol. 7061, SPIE, 2008, pp. 65—75.

[20] M. Jang, Y. Horie, A. Shibukawa, et al., "Wavefront shaping with disorder-engineered metasurfaces," *Nat. Photonics*, vol. 12, no. 2, pp. 84—90, 2018.

[21] E. Tseng, S. Colburn, J. Whitehead, et al., "Neural nano-optics for high-quality thin lens imaging," *Nat. Commun.*, vol. 12, no. 1, p. 6493, 2021.

[22] J. Engelberg and U. Levy, "The advantages of metalenses over diffractive lenses," *Nat. Commun.*, vol. 11, no. 1, p. 1991, 2020.

[23] D. Lin, P. Fan, E. Hasman, and M. L. Brongersma, "Dielectric gradient metasurface optical elements," *Science*, vol. 345, no. 6194, pp. 298—302, 2014.

[24] J. N. Mait, R. A. Athale, J. van der Gracht, and G. W. Euliss, "Potential applications of metamaterials to computational imaging," in *Frontiers in Optics*, Optical Society of America, 2020, p. FTu8B-1.

[25] Y. Peng, Q. Sun, X. Dun, G. Wetzstein, W. Heidrich, and F. Heide, "Learned large field-of-view imaging with thin-plate optics," *ACM Trans. Graph.*, vol. 38, no. 6, pp. 219—221, 2019.

[26] N. Yu and F. Capasso, "Flat optics with designer metasurfaces," *Nat. Mater.*, vol. 13, no. 2, pp. 139—150, 2014.

[27] Z. Lin, C. Roques-Carmes, R. Pestourie, M. Soljačić, A. Majumdar, and S. G. Johnson, "End-to-end nanophotonic inverse design for imaging and polarimetry," *Nanophotonics*, vol. 10, no. 3, pp. 1177—1187, 2020.

[28] A. Arbabi, Y. Horie, M. Bagheri, and A. Faraon, "Dielectric metasurfaces for complete control of phase and polarization with subwavelength spatial resolution and high transmission," *Nat. Nanotechnol.*, vol. 10, no. 11, pp. 937—943, 2015.

[29] G. Zheng, H. Mühlenbernd, M. Kenney, G. Li, T. Zentgraf, and S. Zhang, "Metasurface holograms reaching 80% efficiency," *Nat. Nanotechnol.*, vol. 10, no. 4, pp. 308—312, 2015.

[30] E. Bayati, R. Pestourie, S. Colburn, Z. Lin, S. G. Johnson, and A. Majumdar, "Inverse designed metalenses with extended depth of focus," *ACS Photonics*, vol. 7, no. 4, pp. 873—878, 2020.

[31] A. Zhan, S. Colburn, C. M. Dodson, and A. Majumdar, "Metasurface freeform nanophotonics," *Sci. Rep.*, vol. 7, no. 1, pp. 1—9, 2017.

[32] S. Colburn, A. Zhan, and A. Majumdar, "Metasurface optics for full-color computational imaging," *Sci. Adv.*, vol. 4, no. 2, p. eaar2114, 2018.

[33] L. Huang, J. Whitehead, S. Colburn, and A. Majumdar, "Design and analysis of extended depth of focus metalenses for achromatic computational imaging," *Photonics Res.*, vol. 8, no. 10, pp. 1613—1623, 2020.

[34] Y. Lei, Q. Zhang, Y. Guo, et al., "Snapshot multi-dimensional computational imaging through a liquid crystal diffuser," *Photonics Res.*, vol. 11, no. 3, pp. B111—B124, 2023.

[35] H. Gao, X. Fan, Y. Wang, et al., "Multi-foci metalens for spectra and polarization ellipticity recognition and reconstruction," *Opto-Electron. Sci.*, vol. 2, no. 3, pp. 220026—220031, 2023.

[36] M. Khorasaninejad, A. Y. Zhu, C. Roques-Carmes, et al., "Polarization-insensitive metalenses at visible wavelengths," *Nano Lett.*, vol. 16, no. 11, pp. 7229—7234, 2016.

[37] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," *ArXiv* abs/1505.04597, 2015.

[38] S. Tan, F. Yang, V. Boominathan, A. Veeraraghavan, and G. V. Naik, "3D imaging using extreme dispersion in optical metasurfaces," *ACS Photonics*, vol. 8, no. 5, pp. 1421—1429, 2021.

[39] J. Page Vizcaíno, F. Saltarin, Y. Belyaev, R. Lyck, T. Lasser, and P. Favaro, "Learning to reconstruct confocal microscopy stacks from single light field images," *IEEE Trans. Comput. Imaging*, vol. 7, pp. 775—788, 2021.

[40] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," *CoRR* abs/1412.6980, 2014.

[41] J. Madhusoodanan, "Smart microscopes spot fleeting biology," *Nature*, vol. 614, pp. 378—380, 2023.