optica

Learned rotationally symmetric diffractive achromat for full-spectrum computational imaging

XIONG DUN,[1,3] HAYATO IKOMA,[2] GORDON WETZSTEIN,[2] ZHANSHAN WANG,[1,3] XINBIN CHENG,[1,3,4] AND YIFAN PENG[2,5]

[1] *Institute of Precision Optical Engineering, School of Physics Science and Engineering, Tongji University, Shanghai 200092, China*
[2] *Electrical Engineering Department, Stanford University, Stanford, California 94305, USA*
[3] *MOE Key Laboratory of Advanced Micro-Structured Materials, Shanghai 200092, China*
[4] *e-mail: chengxb@tongji.edu.cn*
[5] *e-mail: evanpeng@stanford.edu*

Received 6 April 2020; revised 25 June 2020; accepted 29 June 2020 (Doc. ID 394413); published 3 August 2020

Diffractive achromats (DAs) promise ultra-thin and light-weight form factors for full-color computational imaging systems. However, designing DAs with the optimal optical transfer function (OTF) distribution suitable for image reconstruction algorithms has been a difficult challenge. Emerging end-to-end optimization paradigms of diffractive optics and processing algorithms have achieved impressive results, but these approaches require immense computational resources and solve non-convex inverse problems with millions of parameters. Here, we propose a learned rotational symmetric DA design using a concentric ring decomposition that reduces the computational complexity and memory requirements by one order of magnitude compared with conventional end-to-end optimization procedures, which simplifies the optimization significantly. With this approach, we realize the joint learning of a DA with an aperture size of 8 mm and an image recovery neural network, i.e., Res-Unet, in an end-to-end manner across the full visible spectrum (429–699 nm). The peak signal-to-noise ratio of the recovered images of our learned DA is 1.3 dB higher than that of DAs designed by conventional sequential approaches. This is because the learned DA exhibits higher amplitudes of the OTF at high frequencies over the full spectrum. We fabricate the learned DA using imprinting lithography. Experiments show that it resolves both fine details and color fidelity of diverse real-world scenes under natural illumination. The proposed design paradigm paves the way for incorporating DAs for thinner, lighter, and more compact full-spectrum imaging systems.

https://doi.org/10.1364/OPTICA.394413

1. INTRODUCTION

High-quality imaging with reduced optical complexity has drawn much interest in both academic and industrial research and development [1,2]. Accordingly, computational imaging, in which much of the aberration correction is shifted from optics to post-processing algorithms [3–8], has been intensively investigated. A detailed review of the use of recent deep learning advances in computational imaging can be found in Ref. [9]. Similarly, the complexity of optics can be further simplified by introducing diffractive optical elements (DOEs) [10–15]. Their advantages of compact form factor, a large and flexible design space, and relatively good off-axis imaging behavior are highly valuable. Integrating diffractive optics [16–19] or even metasurfaces [20,21] in computational imaging has led to many ultra-thin and lightweight camera designs.

Full-spectrum imaging, ubiquitous in the modern sensor system, was traditionally thought too difficult to be realized using single DOEs because their inherently strong chromatic aberrations could lead to very low amplitudes of the optical transfer function (OTF) at almost all wavelengths. Recently, pioneering works using a sequential design approach have been attempted to realize high-quality full-spectrum imaging with optimized DOE [22,23]. In these, a diffractive achromat (DA) is first designed by enforcing the DOE to produce the desired nearly uniform intensity distribution at the focal plane for each wavelength and then removing the wavelength-invariant residual aberrations (WIRAs) via a subsequent image processing step. This sequential design pipeline provides much better full-spectrum imaging quality than that of conventional DOEs (e.g., Fresnel lenses). This is because mitigating the chromatic aberration improves the amplitudes of their OTFs over the full spectrum. However, it may not be the optimal design paradigm from the perspective of computational imaging.

Notably, the optimal OTF distribution of DOEs for full-spectrum computational imaging remains unclear. For sequential design approaches, the desired optimization target may not be the optimal. The ambiguity and complexity of the resulting WIRAs may drastically increase for a DA with a large aperture and a number of achromatic wavelengths [12]. These factors often lead

2334-2536/20/080913-10 Journal © 2020 Optical Society of America

to a sub-optimal DA and further result in image recovery algorithms failing to resolve a high-fidelity image (e.g., loss of details or occurrence of artifacts).

On the other hand, emerging end-to-end design approaches [24–28] that can jointly design the optics and image processing subject to application-domain-specific imaging tasks may provide a better design paradigm for full-spectrum computational imaging using DOEs. The elimination of chromatic aberrations and the resulting WIRAs can be comprehensively tackled through the end-to-end framework. That being said, the resulting WIRAs can certainly adapt well to the utilized image recovery algorithm, and higher full-spectrum image quality can be achieved without the prior knowledge of the optimal OTF distribution.

However, designing practical DAs that have an aperture size of several millimeters, a focal length of several tens of millimeters, and several tens of numbers of achromatic wavelengths has not be achieved using the end-to-end design framework due to the non-trivial computational memory requirement and complex non-convex optimization. Leveraging a rotationally symmetric model can partially solve this problem [29], but existing work uses the rotational symmetry only to reduce the complexity of the non-convex optimization problem while memory requirements remain exorbitant. For instance, the required memory is approximately 20 GB for the forward and backward propagation of a DA with an aperture size of 8 mm over 29 wavelengths [22], which is impractical for most consumer-level GPUs today. Accordingly, state-of-the-art DOE considers only small aperture sized DAs with three wavelengths [24], which is insufficient to guarantee high-fidelity full-spectrum imaging because of the metamerism problem [22], when coupled with RGB sensors.

In this work, we seek to overcome the challenging requirements in computational and memory complexity of end-to-end optimization paradigms and apply it to learn a DA for high-fidelity full-spectrum imaging. Specifically, a rotationally symmetric imaging model is proposed with concentric ring decomposition. This novel rotationally symmetric imaging model, in tandem with an energy regularization term, contributes to reducing the memory consumption and simplifying the non-convex optimization. Further, a deep neural network, i.e., Res-Unet, is applied as the image recovery module, offering the powerful capability of resolving high-fidelity information. We demonstrate our proposed end-to-end designed DA imaging both in simulation and on prototype lenses over the full visible spectrum. In addition, we reveal that the optimal OTF distribution of DOEs for full-spectrum computational imaging is the one exhibiting high amplitudes at high frequencies over the full spectrum as much as possible.

2. END-TO-END LEARNING OF DIFFRACTIVE ACHROMAT AND IMAGE RECOVERY

The proposed end-to-end paradigm jointly learns the parameters of the optics and image recovery algorithm by building a differentiable pipeline architecture consisting of an imaging model, image recovery neural network, and loss, as shown in Fig. 1. Existing end-to-end frameworks suffer from the bottleneck of substantial computational memory requirements and complex non-convex optimization due to two key features: First, the two-dimensional calculation of the point spread function (PSF) of DOEs in their imaging model results in massive computational complexity. Second, the size of the PSF of DOEs is always very large, leading to a large modeled sensor size. In the following, we overcome this bottleneck by introducing the rotationally symmetric parameterization with concentric ring decomposition to the imaging model and an energy regularization in the loss function. Then, we describe the image recovery neural network and an implementation example.

A. Rotationally Symmetric Imaging Model with Concentric Ring Decomposition

The imaging model consists of generating PSFs of the DA and simulating the captured images with these PSFs. Without loss of generality, we assume that the PSF of a DOE is nearly shift invariant within a limited field of view (FOV), which means that on-axis aberration is the dominating factor that degrades the imaging quality. We observe the insight that the ideal shape for a lens, i.e., without on-axis aberration, is inherently rotationally symmetric. Accordingly, we derive a rotationally symmetric imaging model, which drastically simplifies the computational memory requirement and complexity of the non-convex optimization. This is achieved by decomposing the rotationally symmetric DOE to a 1D sum of series of **circ** functions, in which each individual PSF can be represented by the 1D first-order Bessel function of the first kind.

According to the scalar diffraction theory [30], the on-axis PSF $\mathrm{PSF}(x, y, \lambda)$ of a DA in Cartesian coordinates is given as

$$\mathrm{PSF}(x, y, \lambda) = \left| \frac{1}{\lambda f} e^{\frac{ik}{2f}(x^2+y^2)} \iint P(s, t, \lambda) e^{\frac{ik}{2f}(s^2+t^2)} \times e^{-\frac{ik}{f}(xs+yt)} \mathrm{d}s\,\mathrm{d}t \right|^2, \tag{1}$$

where (s, t) and (x, y) are the spatial coordinates at the DA and the sensor (or image) planes, respectively, f is the distance between the lens and the sensor, which is equivalent to the focal length of the lens when the object point is far away, and $P(s, t, \lambda) = A(s, t) e^{ik(n(\lambda)-1)h(s,t)}$ is the complex transmittance function of the DA. λ is the wavelength, $k = \frac{2\pi}{\lambda}$ is the wave number, $n(\lambda)$ is the refractive index of the substrate, $h(s, t)$ is the height map of the DA, and $A(s, t)$ is a **circ** function that represents the aperture of DA.

Using Eq. (1) directly can lead to high computational complexity, e.g., the optimization and calculation grids are ∼16 million in size for the PSF simulation of a DA with an aperture diameter of 8 mm and a feature size of 2 μm. This inevitably leads to substantial memory consumption and optimization difficulties. Reducing the calculation dimension can simplify the computational complexity. Existing end-to-end frameworks have used an unconstrained height map of the DOE, which makes the reduction in calculation dimension unfeasible, as shown in Fig. 2(a).

Instead, as shown in Fig. 2(b), by applying the rotationally symmetric parameterization on the height map of DA, we reduce the number of optimization parameters to 1D and further simplify the complexity of the non-convex optimization. Still, the PSF is computed in 2D. Considering that a practical rotationally symmetric DA should be discretized as a number of concentric rings with a width d, we decompose $P(s, t, \lambda)$ and the additional phase term $e^{\frac{ik}{2f}(s^2+t^2)}$ to a 1D sum of series of **circ** functions [see Fig. 2(c)], expressed as

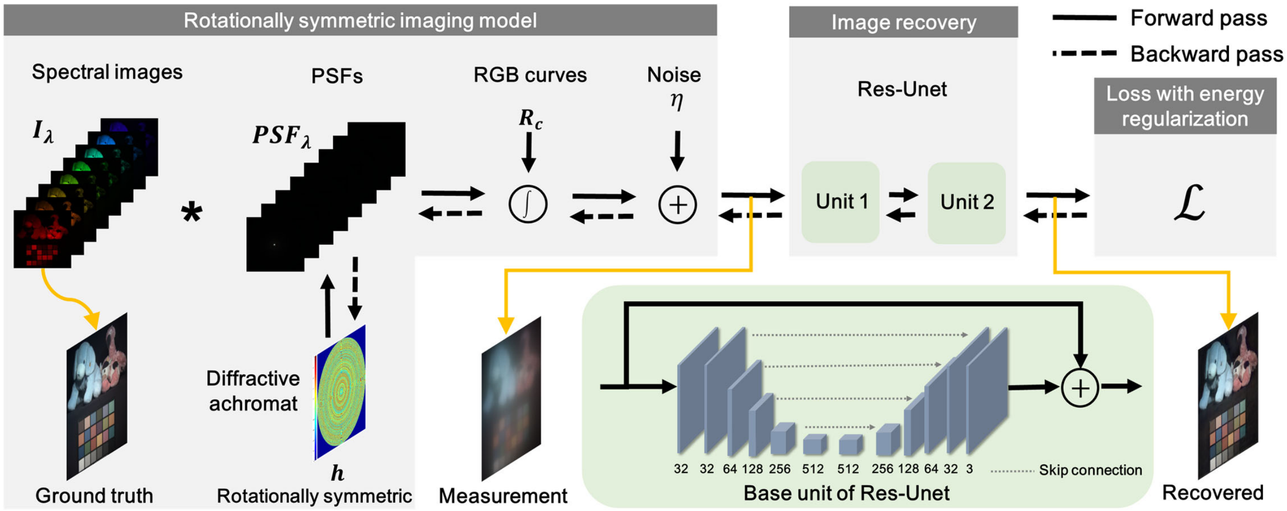

Fig. 1. Overview of proposed end-to-end learning. The parameters of the diffractive achromat (DA) and image recovery algorithm are learned jointly using the end-to-end optimization paradigm. In each forward pass, the spectrally varying scene is convolved with the spectrally varying PSFs of the rotationally symmetric parametrized DA. Then, Gaussian noise is added to the simulated sensor image after integrating over the color response of the RGB sensor for each channel. A neural network, e.g., a Res-Unet consisting of two base network units, is applied as the image recovery unit to resolve a high-fidelity color image. Finally, a differentiable loss, such as the mean squared error for the ground truth image, is defined on the recovered image. An extra energy regularization is added to force light rays to hit within the designated sensor area. In the backward pass, the error is backpropagated to the learned parameters of the image recovery network and height profile of the DA.

$$P(r,\lambda)e^{\frac{ik}{2f}r^2} \approx P(r_1,\lambda)e^{\frac{ik}{2f}r_1^2}\operatorname{circ}\left(\frac{r}{r_1}\right)+\sum_{m=2}^{\infty}P(r_m,\lambda)$$

$$\times\, e^{\frac{ik}{2f}r_m^2}\left[\operatorname{circ}\left(\frac{r}{r_m}\right)-\operatorname{circ}\left(\frac{r}{r_{m-1}}\right)\right], \qquad (2)$$

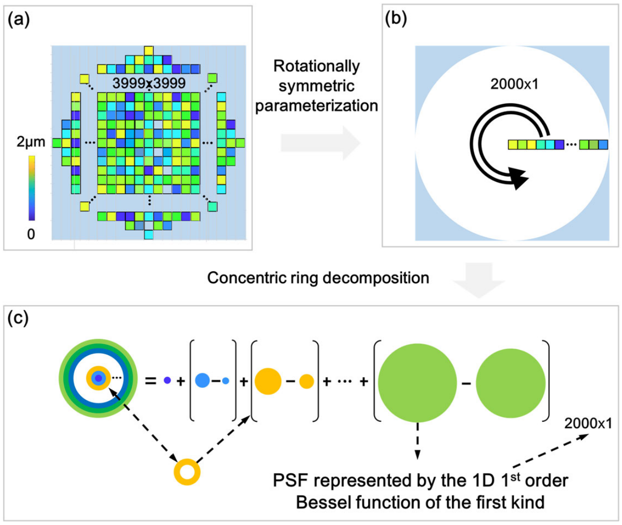
(a)
3999x3999
2μm
0
Rotationally symmetric parameterization
(b)
2000x1
Concentric ring decomposition
(c)
2000x1
PSF represented by the 1D 1st order Bessel function of the first kind

Fig. 2. Principle illustration of the rotationally symmetric imaging model. (a) DOE parameterization in traditional 2D manners is used as a reference. (b) The dimension of optimization parameters can be shrunk to 1D by applying the rotationally symmetric parameterization. (c) The complex transmittance function of the rotationally symmetric DOE is superimposed with a sequence of discrete concentric rings, that can be further decomposed to a 1D sum of series of **circ** functions. Each PSF of **circ** function can be represented by the 1D order Bessel function of the first kind. Using this rotationally symmetric imaging model, the calculation dimension of PSFs and that of optimization parameters can both be reduced to 1D.

where $r=\sqrt{s^2+t^2}$, $r_m=md$, $m=1,2,\ldots$, and circ is the unit **circ** function. For brevity, we derive this in polar coordinates. The approximation of Eq. (2) is reasonable when the ring sampling d is sufficiently fine to accurately approximate the additional phase term $e^{\frac{ik}{2f}r^2}$ by $e^{\frac{ik}{2f}r_m^2}$, for instance, $d \le \frac{\lambda}{2\mathrm{NA}}$ [12], where NA is the numerical number of the DA.

Then, by substituting Eq. (2) into Eq. (1), the rotationally symmetric PSF model of discretized DOEs is derived as follows:

$$\mathrm{PSF}(\rho,\lambda)=\left|\frac{2\pi}{\lambda f}e^{i\frac{k}{2f}(\lambda f\rho)^2}\sum_{m=1}^{\infty}P(r_m,\lambda)e^{\frac{ik}{2f}r_m^2}H(r_m,\rho)\right|^2, \qquad (3)$$

where $\rho=\frac{\sqrt{x^2+y^2}}{\lambda f}$, and $H(r_m,\rho)$ is defined as

$$H(r_m,\rho)=\begin{cases}\frac{1}{2\pi\rho}[r_m J_1(2\pi\rho r_1)-r_{m-1}J_1(2\pi\rho r_{m-1})], & m>1\\ \frac{1}{2\pi\rho}r_1 J_1(2\pi\rho r_1), & m=1\end{cases}, \qquad (4)$$

where J_1 is the first-order Bessel function of the first kind. Please refer to Supplement 1 for derivation details. Note that $H(r_m,\rho)$ can be pre-calculated, and then Eq. (3) can be implemented by single vector-matrix multiplications.

We note that alternative parameterization models, such as the circularly symmetric truncated Zernike base [31], may not work appropriately for end-to-end designs of a DOE lens considering its continuous surface representation. Please refer to Supplement 1 for a detailed evaluation using the Zernike base representation.

Finally, the dimension of the PSF is restored by sweeping the $\mathrm{PSF}(\rho,\lambda)$ around the optical axis of the DA. Then, the images $Y_c(x,y)$ captured by the DA and sensor can be simulated as follows:

$$Y_c(x,y)=\int_{\lambda_{\min}}^{\lambda_{\max}}[PSF(x,y,\lambda)*I(x,y,\lambda)]R_c(\lambda)\mathrm{d}\lambda+\eta, \qquad (5)$$

where $[\lambda_{\min}, \lambda_{\max}]$ is the spectrum range, and η is the sensor read noise (Gaussian noise $\eta \sim \mathcal{N}(0, \sigma^2)$).

Again, this rotationally symmetric imaging model simplifies the calculation of PSFs from 2D to 1D, which reduces the memory consumption and computational complexity by an order of magnitude. As such, for the first time, we can fit a larger lens model (i.e., larger pixel counts and more wavelengths) to commercial GPUs. Akin to conventional rotationally symmetric designs, our approach also leads to a more robust optimization of DAs by reducing the number of variables, e.g., for the design shown in Section 2.D, the number of variables is reduced from 16,000,000 to 2000.

B. Loss with Energy Regularization

The outer diameter of the PSF of the DA is determined by its feature size, focal length, and wavelength. For instance, at the wavelength of 550 nm, the PSF diameter goes up to 13.75 mm (calculated by $\frac{\lambda f}{d}$) when the DA feature size and its focal length are 2 µm and 50 mm, respectively. That said, the modeled sensor size in the design space should be larger than 13.75 mm to guarantee accurate modeling of the PSF. However, this results in large-sized patches of the input data (e.g., larger than 2,292 × 2,292 when the sensor pixel size is set to 6 µm), and inevitably increases the memory consumption, thereby further hindering the optimization from being implemented on commercial GPUs.

We introduce an energy regularization term, forcing most of the light rays to hit the designated sensor area. This setting allows us to use a relatively small sensor size for saving device memory. This small sensor size may cause a severe deviation of the synthetic forward imaging pass in Fig. 1 from the realistic forward imaging pass. However, the energy regularization contributes to bridging this gap. Specifically, we penalize a regularization term $\mathcal{R}(\mathrm{PSF})$ that calculates the energy of light rays missing the designated sensor area, which is defined as

$$\mathcal{R}(\mathrm{PSF}) = \int_{\lambda_{\min}}^{\lambda_{\max}} \iint W(x, y)\mathrm{PSF}(x, y, \lambda)\mathrm{d}x\mathrm{d}y\mathrm{d}\lambda, \tag{6}$$

where $W(x, y)$ is a selecting mask for indexing those pixels that fall outside the modeled sensor area by setting the pixel value to one, and otherwise to zero.

The loss function of the proposed framework combines this energy regularization as well as the ℓ_2 (mean squared error) loss, which evaluate the errors between the recovery and original ground truth [our approach can also generalize to alternative data fidelity losses, such as ℓ_1 loss or structural similarity index (SSIM) loss], expressed as

$$\mathcal{L} = \sum_c ||\widetilde{G_c}(x, y) - G_c(x, y)||_2^2 + \alpha\mathcal{R}(\mathrm{PSF}), \tag{7}$$

where α is the regularization weight, $\widetilde{G_c}$ is the output of Res-Unet, and G_c is the ground truth RGB image determined by the RGB spectral response curve, $R_c(\lambda)$, of the sensor, represented as

$$G_c(x, y) = \int_{\lambda_{\min}}^{\lambda_{\max}} I(x, y, \lambda)R_c(\lambda)\mathrm{d}\lambda. \tag{8}$$

We evaluate the energy-preserving behavior with and without the energy regularization under the design parameters shown in Section 2.D. As shown in Fig. 3, although the PSF is slightly sharper without the energy regularization, a considerable amount of energy falls outside the designated sensor area (only ∼6% is measured by the designated sensor). This would result in an ill-conditioned end-to-end optimization problem because of the severely truncated PSFs used in the forward imaging pass. In contrast, the energy of the averaged PSF falling within the designated sensor area increased to ∼76% after adding the energy regularization. This observation agrees well with our argument above that the energy regularization contributes to bridging the gap between the synthetic and realistic forward imaging pass.

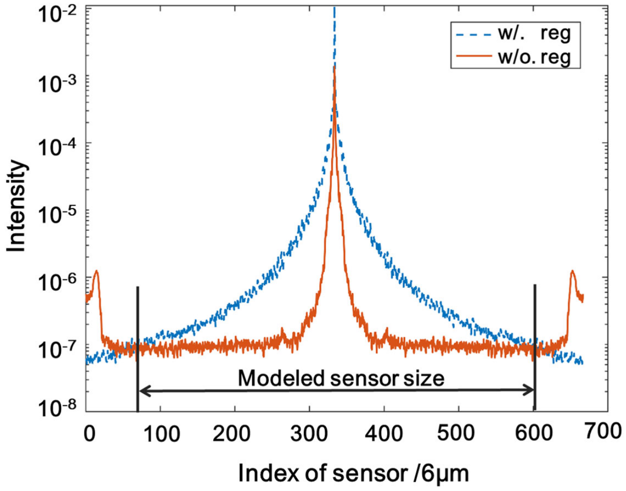

Fig. 3. PSFs of the DA designed with and without energy regularization. We present the cross section of the 2D PSFs that are normalized to the input energy.

Note that adding the energy regularization term and using the rotationally symmetric imaging model, the memory consumption reduces from ∼20GB to ∼2GB for the forward and backward propagation of a DA with an aperture size of 8 mm over 31 wavelengths.

C. Image Recovery Neural Network

To offer the powerful capability of resolving high-fidelity information from the measurements degraded by the WIRAs of the DA and the sensor noise, and inspired by the recent success of deep image recovery networks [7,32–35], we develop a residual-Unet (Res-Unet) based on the original Unet [36], which implements the multi-scale operation on the image, and Res-net [37], which can extend the network to a much deeper architecture for augmenting performance. Incorporating these two networks imparts several advantages. First, it performs a multi-scale operation on the image because of the contracting and expanding paths, making it suitable for the image recovery of scenarios where large blur occurs; second, it learns the residual image, i.e., the difference between the sharp latent image and the measurement, instead of learning the original sharp image itself, making the optimization of the network easier; third, it scales to a deeper and more robust architecture by directly extending the network with a base unit of the Res-Unet.

The architecture of the base unit of the Res-Unet is presented in the bottom right corner of Fig. 1, where in our current implementation, two base units are applied. In relevant work, two consecutive U-net blocks can also be referred as W-net [38]. The first block of the unit consists of a convolution layer with 32 kernels of a 3 × 3 matrix. Each of the second to sixth blocks of the unit consists of a convolution layer with 32 kernels of a 4 × 4 matrix, and a stride of two for the downsampling; we double the number of kernels at each downsampling convolution. Each of the seventh to eleventh blocks of the unit consists of a nearest neighbor upsampling and a 3 × 3 convolution layer. Moreover, we concatenate the outputs from the downsampling layer to introduce high frequencies to preserve fine scene details. A final convolution layer with three kernels of a 3 × 3 matrix is used to make the residual output size equal to the original image size. The residual image is added to the sensor output, and through a clamp layer, which keeps the final output data between zero and one. All blocks but the last one include a leaky rectified linear unit (L-ReLU with slope = 0.2) as the activation function. Please refer to Supplement 1 for the configuration details.

D. Implementation Example

We implement our proposed framework with TensorFlow [39] and optimize with stochastic gradient methods. The optimization variables include the rotationally symmetric height profile, $h(r)$, and the parameters of the convolutional kernels of the Res-Unet. The feature size and focal length of the DA, and the sensor pixel size and read noise level are hyper-parameters. We experimentally observe that the Adam optimizer works appropriately for this joint design framework.

Specifically, we design an achromatic diffractive lens with a focal length of 50 mm and an aperture diameter of 8 mm. The feature size of the DOE is set to 2 µm, corresponding to a total pixel number in the design of 4000 × 4000. The achromatic spectrum range is from 429 to 699 nm. The material of the designed DOE is NOA61.

The dataset used for training is built on the datasets made public by Harvard [40], ICVL [41], CAVE [41], and NUS [42]. We have a total of 376 hyperspectral images. Among them, 15 images are used as the test set, while the remaining 361 images are for training. We have 361 iterations in one epoch. In each iteration of the training epochs, we sequentially read in one hyperspectral image and randomly crop it into patches, each with a pixel count of 512 × 512. Then, we randomly flip and rotate all patches to augment the training process.

To contribute to a relatively smooth surface profile for a robust fabrication process using grayscale photolithography, we introduce a filtering process into the optimization framework to locally smooth the height profile. Specifically, we extract the height profile after 10 epochs, and implement the median filtering with a kernel size of five on the current height profile. Then, we feed the filtered height profile back and continue the training process. In our current training implementation, this filtering process is re-implemented every 10 epochs. We also add a random uniform noise distribution in the range ±20 nm to the height profile before computing the PSF to augment the robustness of fabrication imperfections.

We model the sensor with a pixel count of 512 × 512 and a pixel size of 6 µm. The sensor read noise is Gaussian noise with a standard deviation drawn from a uniform distribution between 0.001 and 0.015 (with an image scale of [0,1]). Each hyperspectral image contains 31 channels representing the range from 429 to 699 nm with an interval of 9 nm. The regularization weight, α, is set to 1e-4. The height profile of the DOE is initialized to 10 nm at the beginning of the optimization. We use a learning rate of 10^{-4} with the Adam optimizer. The optimization phase is run for 215 epochs, which takes approximately 18 h on a single NVIDIA 1080Ti GPU. We experimentally observe that using the ℓ_2 loss leads to better results, e.g., in higher peak signal-to-noise ratio (PSNR), than ℓ_1 loss and SSIM loss in our case. Please refer to Supplement 1 for more details.

3. ASSESSMENT IN SIMULATION

We first assess the final imaging performance of *three* diffractive lenses: (1) a conventional Fresnel lens that focuses light at a single wavelength (555 nm), which is used as the baseline of regular DOEs showing strong chromatic aberration; (2) a reference DA with the same design parameters as in Section 2.D, designed in a conventional sequential manner following the scheme described in a prior work [22]; and (3) the proposed DA designed using our end-to-end paradigm.

For a fair comparison, we use the same training set and parameters to train the image recovery Res-Unets for the Fresnel lens and the reference DAs. We assess 15 test images using the objective criteria of the PSNR, SSIM, and spectral angular mapper (SAM) [43]. Generally, a higher PSNR, a higher SSIM, or a lower SAM, indicates a closer agreement between the recovered images and ground truth data. Gaussian noise with $\sigma = 0.003$ is added to all sensor measurements. Table 1 summarizes the average PSNR, SSIM, and SAM, and Fig. 4 shows several examples selected from the test set. We observe that the Fresnel lens exhibits significantly worse performance than ours in PSNR, SSIM, SAM, and visual performance, as shown in Fig. 4. Concerning the reference DA, although the performance is much better than that of the Fresnel lens in both spatial and spectral quality (with averaged PSNR and SAM improvements of 6 dB and 0.04, respectively), it resolves fewer details [see Fig. 4(a)] and suffers from more artifacts [see Fig. 4(b)] than that of our end-to-end designed DA. The averaged PSNR, SSIM, and SAM of the reference DA are = 1.3 dB, 0.015, and 0.01, respectively, lower than those of the proposed one. As such, we demonstrate that the proposed end-to-end optimization framework can lead to superior results when deriving DAs for full-spectrum applications. Results with noise levels $\sigma = 0.006$ and $\sigma = 0.011$ are presented in Supplement 1.

To explore why the end-to-end design leads to better results than that of conventional sequential methods, we assess the focusing behavior of two DAs. The result of a Fresnel lens is also shown as the reference. Figures 5(a), 5(d), and 5(g) show a stack of the focused light intensity profiles of a Fresnel lens, the proposed DA, and the reference DA, respectively. A Fresnel lens can focus only one single wavelength at the designed focal distance. However, both DAs have the identical focal plane at all 31 designed wavelengths, showing the achromatic behavior over the full spectrum We also show five selected PSFs derived subject to the spectrum of different color targets in Figs. 5(b), 5(e), and 5(h). We observe three insights: (1) the Fresnel lens suffers from strong chromatic aberration; (2) the WIRAs are inevitable when aiming to mitigate the chromatic aberration of DOEs; and (3) the WIRAs vary subject to design methods.

Next, to further investigate why the WIRAs of the proposed DA are optimal for the subsequent image recovery algorithms and what is the optimal OTF distribution for full-spectrum computational imaging with DOEs, we explore the OTF of the three lenses in Figs. 5(c), 5(f), and 5(i). We observe that the amplitudes of the OTF of the Fresnel lens at high frequencies are pretty low (~0.0005) at most wavelengths. Instead, the amplitudes of the OTF of the reference and proposed DAs at high frequencies are

Table 1. Quantitative Evaluation of Averaged PSNR (dB), SSIM, and SAM over 15 Test Images Resolved Using Different Lens Designs and Recovery Algorithms

Lens Design	Fresnel Lens	Reference DA	Proposed DA
Measurement	19.43/0.645/0.17	19.80/**0.662**/0.15	**19.90**/0.657/**0.14**
Recovery	25.78/0.804/0.12	31.79/0.877/0.08	**33.09/0.892/0.07**

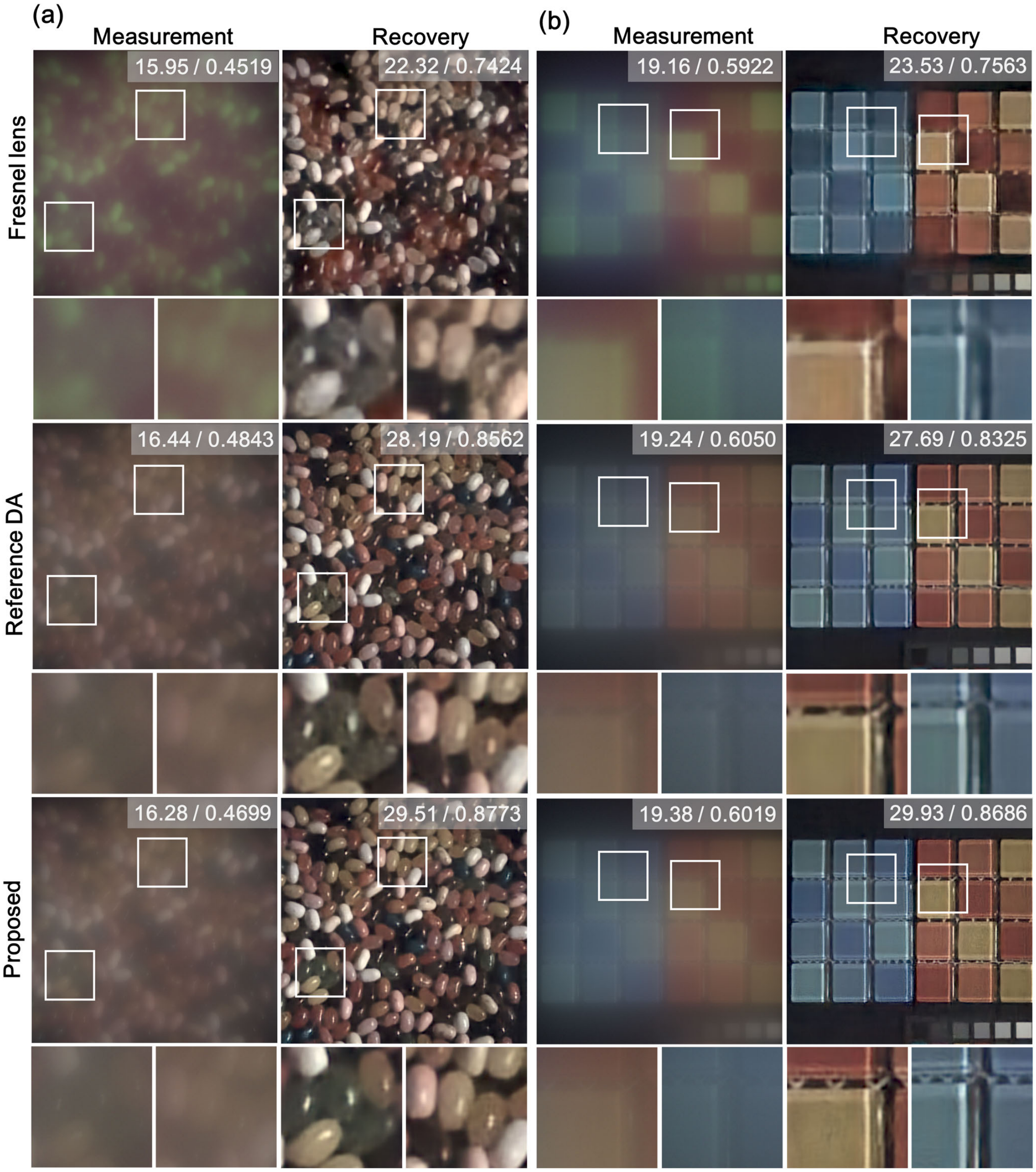

Fig. 4. Selected examples of the assessment of the DA designs in simulation. We assess the performance of a conventional Fresnel lens, a reference DA, and a DA optimized using the proposed framework. We show both the original sensor measurements and the recovery results of the Res-Unet. The inset values indicate the PSNR (dB) and SSIM.

~0.01 and ~0.025, respectively. This observation suggests that the amplitude of the OTF at high frequencies matters, concerning improving the performance of diffractive full-spectrum computational imaging. Intuitively, the higher the amplitude of the OTF in high-frequency range, the higher quality of the recovery image. Although the sequential design manner has already led to a higher amplitude of the OTF at high frequencies than that of a Fresnel lens, the resulting DA design is often sub-optimal. This is because the amplitudes of OTFs of specified target PSFs are always lower at high frequencies than those at low and mid frequencies. As such, the non-convex optimization may fail to enforce an increase of the amplitude of the OTF at high frequencies, especially for cases

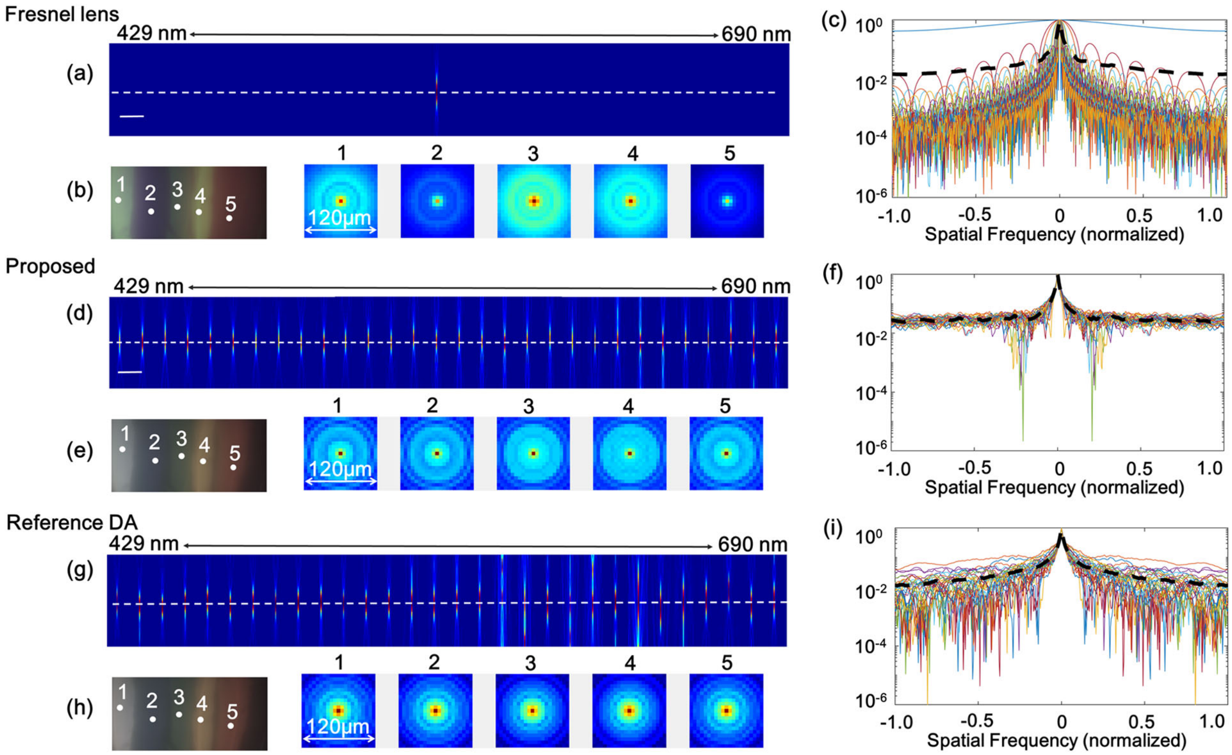

Fig. 5. Simulated performance of a Fresnel lens and two DAs. For each design, we show a stack of the focused light intensity profiles (a), (d), (g) along the optical axis at multiple wavelengths, where the white dashed line indicates the focal plane, and the scale bar corresponds to 200 µm. We also show the simulated sensor measurements and the normalized PSFs (b), (e), (h) of selected scene points. The normalized PSFs are shown in log scale for visualization purpose. In addition, we show the OTFs of three lenses (c), (f), (i) at 31 design wavelengths, respectively. The black dashed line shows the averaged OTF of the 31 design wavelengths.

where the complexity of WIRAs is notable. However, via enforcing the joint optimization, the design space is comprehensively explored with the consideration of final image quality. As such, the WIRAs can adapt to the specific image recovery algorithm and lead to higher amplitudes of the OTF at high frequencies.

4. EXPERIMENTAL RESULTS

A. Prototype

The designed DOEs are fabricated using imprinting lithography [44]. Please refer to Supplement 1 for the details. Figure 6(a) illustrates a microscopy image of the fabricated DOE.

The fabricated lens was then attached to a Canon T5i DSLR camera body with 5, 740 × 3, 648 pixels and a pixel pitch of 4.1 µm. To account for deviations between the designed and fabricated DA, we used a white LED light source with a 35 µm pinhole attached in front to calibrate the real-world PSFs of our DA, as shown in Fig. 6(b). The measured PSFs exhibit a slight deviation from the designed ones due to the imperfect fabrication (with a mean square error of 0.07 for the image scale of [0,1]). A fine-tune of the image recovery network with the measured PSF can mitigate the influence of fabrication error. Please refer to Supplement 1 for elaborated analysis.

B. Full Field-of-View Imaging Behavior

Our proposed DA exhibits visually identical on-axis and off-axis performances. As illustrated in Fig. 7(a), the checkerboard is degraded uniformly within the full FOV of 18.6°, demonstrating the assumption that the PSF of the DA is FOV independent. After a post-processing step using the Res-Unet, the entire checkerboard is resolved. We use the modulation transfer function (MTF), one

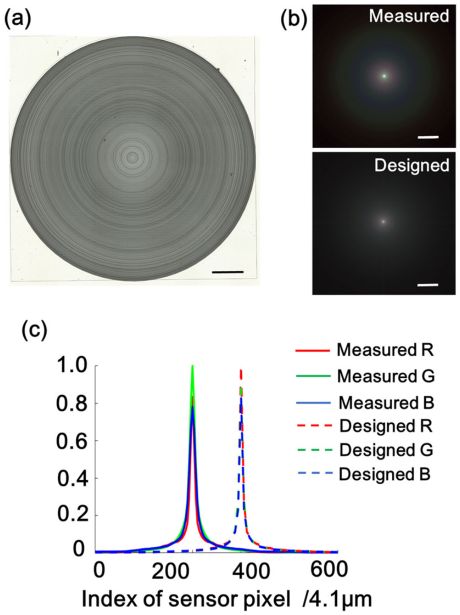

Fig. 6. Measurement of the proposed DA: (a) microscopy image of the fabricated DA (scale bar indicates 0.5 mm); (b) designed and measured PSFs with a Canon T5i DSLR camera body (scale bar indicates 60 µm); and (c) cross section of the PSFs of (b). The PSFs shown here are gamma-corrected for visualization purpose.

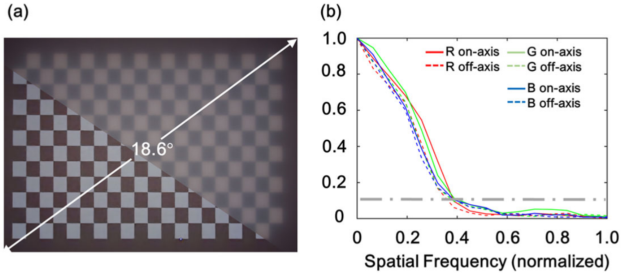

Fig. 7. Evaluation of full field-of-view (FOV) imaging behavior: (a) degraded and recovered checkerboard image pair and (b) MTFs estimated from the grayscale slant edges inside (a); on-axis and off-axis represent the 0° and 17.5° FOVs, respectively.

of the representative metrics in optical designs, and estimate it on the recovered image using the slant edge method [45], as presented in Fig. 7(b). Although in computational imaging, defining MTFs in a conventional manner is not the optimal metric to evaluate spatial resolution, still these plots intuitively reveal the full FOV imaging behavior. We observe a reasonable compromise between the on-axis and off-axis performances as well as an MTF larger than 0.1 at the normalized frequency of 0.35 (corresponding to 85 lp/mm).

C. Captured Results

The experimental results captured using our DA are shown in Fig. 8, presenting diverse real-world scenes including indoor, outdoor, large reflection feature, and rich color, under both artificial illumination and natural light. The results show that our method successfully preserves both fine details and color fidelity. It is clear that our DA combined with the Res-Unet is able to perform high-fidelity full-spectrum imaging. We note that the color of the checkerboard scene deviates from that of a standard checkerboard. It is mainly because this image is displayed on an LCD monitor whose color gamut may deviate from the natural color gamut. More results captured by another machine vision sensor (Pointgrey Grasshopper3 USB3) can be found in Fig. S3 (Supplement 1), indicating that the fabricated DA is achromatic due to its robustness to different spectral response curves of the sensors.

5. CONCLUSION

We have presented a memory-efficient end-to-end design paradigm for full-spectrum computational imaging with diffractive optics. This is achieved by deriving a novel rotationally symmetric PSF model and incorporating the energy regularization in the loss function. Superior high-fidelity imaging performance has been realized by jointly learning a DA and an image recovery neural network. The design paradigm maximizes the amplitudes of OTF at high frequencies over the full spectrum and inherently seeks the optimal solution via comprehensively tackling both chromatic aberration and WIRAs. As such, we have demonstrated realistic DA imaging with a large pixel count (e.g., 4000 × 4000), multi-wavelength channels (e.g., 31), and a sophisticated deep neural network (e.g., Res-Unet). We envision our method to

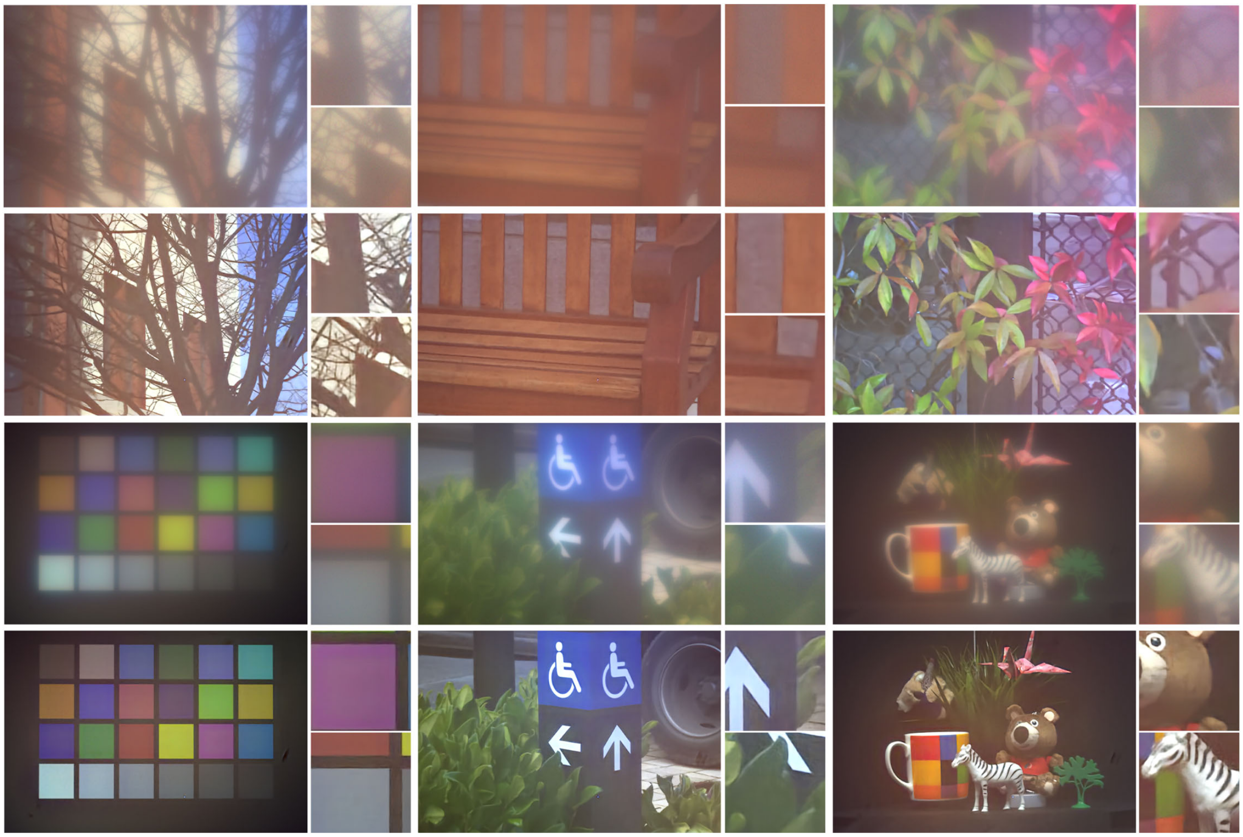

Fig. 8. Experimental results of the fabricated DA. For each pair, we show the degraded sensor measurement and the recovery result. The exposure times for these images are 2.5, 125, 76, 600, 25, and 600 ms (respectively, from left to right, top to bottom) at ISO 100. The images are center-cropped regions (3, 000 × 2, 000) of the original camera measurement. The processing time at this image size on an NVIDIA 1080Ti GPU is around 4 s.

lay a foundation for enabling DAs for thinner, lighter, and more compact full-spectrum cameras, and additionally establish the potential of extending compact computational cameras with DOEs to widespread applications such as single-shot depth imaging, high-dynamic-range imaging, and other high-level vision tasks.

Funding. National Natural Science Foundation of China (61925504, 61621001); Special development fund for Shanghai Zhangjiang National Independent Innovation Zone (ZJ2019-ZD-003); National Key Research and Development Program of China (2016YFA0200900, 2016YFA0200901); Major projects of Science and Technology Commission of Shanghai (17JC1400800); Shu Guang project supported by Shanghai Municipal Education Commission and Shanghai Education (17SG22); Fundamental Research Funds for the Central Universities; National Science Foundation CAREER Award (1839974, 1553333); PECASE by the Army Research Laboratory; National Science Foundation (ECCS-1542152).

Disclosures. The authors declare no conflicts of interest.

See Supplement 1 for supporting content.

REFERENCES

1. D. L. Marks, D. R. Golish, D. J. Brady, D. S. Kittle, E. J. Tremblay, E. M. Vera, S. S. Hui, J. E. Ford, J. Kim, and M. E. Gehm, “Gigapixel imaging with the aware multiscale camera,” Opt. Photon. News **23**(12), 31 (2012).
2. K. Venkataraman, L. Dan, A. Mcmahon, G. Molina, P. Chatterjee, R. Mullis, and S. Nayar, “PiCam: an ultra-thin high performance monolithic camera array,” ACM Trans. Graph. **32**, 1–13 (2013).
3. F. Heide, M. Rouf, M. B. Hullin, B. Labitzke, W. Heidrich, and A. Kolb, “High-quality computational imaging through simple lenses,” ACM Trans. Graph. **32**, 149 (2013).
4. N. Antipa, G. Kuo, R. Heckel, B. Mildenhall, E. Bostan, R. Ng, and L. Waller, “DiffuserCam: lensless single-exposure 3D imaging,” Optica **5**, 1–9 (2018).
5. M. S. Asif, A. Ayremlou, A. Veeraraghavan, R. Baraniuk, and A. Sankaranarayanan, “Flatcam: replacing lenses with masks and computation,” in *IEEE International Conference on Computer Vision (ICCV)* (2015), pp. 663–666.
6. A. Sinha, J. Lee, S. Li, and G. Barbastathis, “Lensless computational imaging through deep learning,” Optica **4**, 1117–1125 (2017).
7. Y. Peng, Q. Sun, X. Dun, G. Wetzstein, and W. Heidrich, “Learned large field-of-view imaging with thin-plate optics,” ACM Trans. Graph. **38**, 219 (2019).
8. K. Monakhova, J. Yurtsever, G. Kuo, N. Antipa, K. Yanny, and L. Waller, “Learned reconstructions for practical mask-based lensless imaging,” Opt. Express **27**, 28075–28090 (2019).
9. G. Barbastathis, A. Ozcan, and G. Situ, “On the use of deep learning for computational imaging,” Optica **6**, 921–943 (2019).
10. P. R. Gill and D. G. Stork, “Lensless ultra-miniature imagers using odd-symmetry spiral phase gratings,” in *Imaging and Applied Optics* (2013), paper CW4C.3.
11. S. Banerji and B. Sensale-Rodriguez, “A computational design framework for efficient, fabrication error-tolerant, planar THz diffractive optical elements,” Sci. Rep. **9**, 5801 (2019).
12. S. Banerji, M. Meem, A. Majumder, F. G. Vasquez, B. Sensale-Rodriguez, and R. Menon, “Imaging with flat optics: metalenses or diffractive lenses?” Optica **6**, 805–810 (2019).
13. M. Meem, S. Banerji, C. Pies, T. Oberbiermann, A. Majumder, B. Sensale-Rodriguez, and R. Menon, “Large-area, high-numerical-aperture multi-level diffractive lens via inverse design,” Optica **7**, 252–253 (2020).
14. S. Banerji, M. Meem, A. Majumder, B. Sensale-Rodriguez, and R. Menon, “Extreme-depth-of-focus imaging with a flat lens,” Optica **7**, 214–217 (2020).
15. P. Wang, N. Mohammad, and R. Menon, “Chromatic aberration corrected diffractive lenses for ultra broadband focusing,” Sci. Rep. **6**, 21545 (2016).
16. Y. Peng, Q. Fu, H. Amata, S. Su, F. Heide, and W. Heidrich, “Computational imaging using lightweight diffractive-refractive optics,” Opt. Express **23**, 31393–31407 (2015).
17. F. Heide, Q. Fu, Y. Peng, and W. Heidrich, “Encoded diffractive optics for full-spectrum computational imaging,” Sci. Rep. **6**, 33543 (2016).
18. D. S. Jeon, S.-H. Baek, S. Yi, Q. Fu, X. Dun, W. Heidrich, and M. H. Kim, “Compact snapshot hyperspectral imaging with diffracted rotation,” ACM Trans. Graph. **38**, 117 (2019).
19. Y. Peng, X. Dun, Q. Sun, F. Heide, and W. Heidrich, “Focal sweep imaging with multi-focal diffractive optics,” in *IEEE International Conference on Computational Photography (ICCP)* (2018), pp. 1–8.
20. S. Colburn, A. Zhan, and A. Majumdar, “Metasurface optics for full-color computational imaging,” Sci. Adv. **4**, eaar2114 (2018).
21. S. Colburn, A. Zhan, and A. Majumdar, “Varifocal zoom imaging with large area focal length adjustable metalenses,” Optica **5**, 825–831 (2018).
22. Y. Peng, F. Qiang, F. Heide, and W. Heidrich, “The diffractive achromat full spectrum computational imaging with diffractive optics,” ACM Trans. Graph. **35**, 31 (2016).
23. N. Mohammad, M. Meem, B. Shen, P. Wang, and R. Menon, “Broadband imaging with one planar diffractive lens,” Sci. Rep. **8**, 2799 (2018).
24. V. Sitzmann, S. Diamond, Y. Peng, X. Dun, S. Boyd, W. Heidrich, F. Heide, and G. Wetzstein, “End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging,” ACM Trans. Graph. **37**, 1–13 (2018).
25. J. Chang and G. Wetzstein, “Deep optics for monocular depth estimation and 3D object detection,” in *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)* (2019), pp. 10193–10202.
26. Y. Wu, V. Boominathan, H. Chen, A. Sankaranarayanan, and A. Veeraraghavan, “PhaseCam3D—learning phase masks for passive single view depth estimation,” in *IEEE International Conference on Computational Photography (ICCP)* (2019), pp. 1–12.
27. C. A. Metzler, H. Ikoma, Y. Peng, and G. Wetzstein, “Deep optics for single-shot high-dynamic-range imaging,” in *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)* (2020).
28. J. Chang, V. Sitzmann, X. Dun, W. Heidrich, and G. Wetzstein, “Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification,” Sci. Rep. **8**, 12324 (2018).
29. H. Haim, S. Elmalem, R. Giryes, A. M. Bronstein, and E. Marom, “Depth estimation from a single image using deep learned phase coded mask,” IEEE Trans. Comput. Imaging **4**, 298–310 (2018).
30. J. W. Goodman, *Introduction to Fourier Optics* (Roberts and Company, 2005).
31. Y. Shechtman, S. J. Sahl, A. S. Backer, and W. E. Moerner, “Optimal point spread function design for 3D imaging,” Phys. Rev. Lett. **113**, 133902 (2014).
32. K. He, X. Zhang, S. Ren, J. Sun, B. Leibe, J. Matas, N. Sebe, and M. Welling, “Identity mappings in deep residual networks,” in *European Conference on Computer Vision (ECCV)* (2016), pp. 630–645.
33. K. Zhang, W. Zuo, S. Gu, and L. Zhang, “Learning deep CNN denoiser prior for image restoration,” in *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), pp. 3929–3938.
34. S. Elmalem, R. Giryes, and E. Marom, “Learned phase coded aperture for the benefit of depth of field extension,” Opt. Express **26**, 15316–15331 (2018).
35. S. Nah, T. Hyun Kim, and K. Mu Lee, “Deep multi-scale convolutional neural network for dynamic scene deblurring,” in *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), pp. 3883–3891.
36. O. Ronneberger, P. Fischer, T. Brox, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, “U-net: convolutional networks for biomedical image segmentation,” in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)* (2015), pp. 234–241.
37. K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)* (2016), pp. 770–778.

38. X. Xia and B. Kulis, “W-net: a deep model for fully unsupervised image segmentation,” arXiv preprint arXiv:1711.08506 (2017).
39. Google, “TensorFlow: large-scale machine learning on heterogeneous systems,” 2015, https://tensorflow.org.
40. A. Chakrabarti and T. Zickler, “Statistics of real-world hyperspectral images,” in *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)* (2011), pp. 193–200.
41. F. Yasuma, T. Mitsunaga, D. Iso, and S. Nayar, “Generalized assorted pixel camera: post-capture control of resolution, dynamic range and spectrum,” Technical Report CUCS-061-08 (Columbia University, 2008).
42. R. M. Nguyen, D. K. Prasad, and M. S. Brown, “Training-based spectral reconstruction from a single RGB image,” in *European Conference on Computer Vision (ECCV)* (2014), pp. 186–201.
43. F. A. Kruse, A. B. Lefkoff, J. W. Boardman, K. B. Heidebrecht, A. T. Shapiro, P. J. Barloon, and A. F. H. Goetz, “The spectral image processing system (SIPS)-interactive visualization and analysis of imaging spectrometer data,” Remote. Sens. Environ. **44**, 145–163 (1993).
44. Y. Xia and G. M. Whitesides, “Soft lithography,” Annu. Rev. Mater. Sci. **28**, 153–184 (1998).
45. E. Samei, M. J. Flynn, and D. A. Reimann, “A method for measuring the presampled MTF of digital radiographic systems using an edge test device,” Med. Phys. **25**, 102–113 (1998).