

## Learned rotationally symmetric diffractive achromat for full-spectrum computational imaging: supplement

**XIONG DUN,<sup>1,3</sup> HAYATO IKOMA,<sup>2</sup> GORDON WETZSTEIN,<sup>2</sup> ZHANSHAN WANG,<sup>1,3</sup> XINBIN CHENG,<sup>1,3,4</sup> AND YIFAN PENG<sup>2,5</sup>**

<sup>1</sup>*Institute of Precision Optical Engineering, School of Physics Science and Engineering, Tongji University, Shanghai 200092, China*

<sup>2</sup>*Electrical Engineering Department, Stanford University, Stanford, California 94305, USA*

<sup>3</sup>*MOE Key Laboratory of Advanced Micro-Structured Materials, Shanghai 200092, China*

<sup>4</sup>*e-mail: chengxb@tongji.edu.cn*

<sup>5</sup>*e-mail: evanpeng@stanford.edu*

---

This supplement published with The Optical Society on 3 August 2020 by The Authors under the terms of the [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/) in the format provided by the authors and unedited. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

Supplement DOI: <https://doi.org/10.6084/m9.figshare.12588728>

Parent Article DOI: <https://doi.org/10.1364/OPTICA.394413>

# Learned Rotationally Symmetric Diffractive Achromat for Full-Spectrum Computational Imaging: Supplementary Information

XIONG DUN<sup>1,3</sup>, HAYATO IKOMA<sup>2</sup>, GORDON WETZSTEIN<sup>2</sup>, ZHANSHAN WANG<sup>1,3</sup>,  
XINBIN CHENG<sup>1,3,4</sup>, AND YIFAN PENG<sup>2,5</sup>

<sup>1</sup>Institute of Precision Optical Engineering, School of Physics Science and Engineering, Tongji University, Shanghai, 200092, China

<sup>2</sup>Electrical Engineering Department, Stanford University, Stanford, CA, 94305, USA

<sup>3</sup>MOE Key Laboratory of Advanced Micro-Structured Materials, Shanghai, 200092, China

<sup>4</sup>chengxb@tongji.edu.cn

<sup>5</sup>evanpeng@stanford.edu

Compiled July 14, 2020

This document provides supplementary information to *Learned Rotationally Symmetric Diffractive Achromat for Full-Spectrum Computational Imaging*. We provide detailed derivation of the rotationally symmetric point spread function (PSF) model. We also provide the additional details about the utilized image recovery neural network and the fabrication. In addition, we provide the comparison of the proposed Res-UNet recovery with model-based methods, that under different loss function configurations, and that designed with the Zernike base parameterization, as well as extra experimental results captured with our fabricated DA but using a machine vision sensor.

## 1. ROTATIONALLY SYMMETRIC PSF MODEL

In the following, we describe the details of deriving Eq. 3 in main text. We first rewrite Eq. 1 in main text as:

$$PSF(x, y, \lambda) = \left| \frac{1}{\lambda f} e^{\frac{ik}{2f}(x^2+y^2)} \int \int P(s, t, \lambda) e^{\frac{ik}{2f}(s^2+t^2)} e^{-\frac{ik}{f}(xs+yt)} ds dt \right|^2. \quad (S1)$$

By substituting  $k = \frac{2\pi}{\lambda}$  and assuming  $f_x = \frac{x}{f\lambda}$  and  $f_y = \frac{y}{f\lambda}$ , Eq. S1 can be expressed as:

$$PSF(f_x, f_y, \lambda) = \left| \frac{1}{\lambda f} e^{i\pi\lambda f(f_x^2+f_y^2)} \int \int P(s, t, \lambda) e^{\frac{ik}{2f}(s^2+t^2)} e^{-i2\pi(f_x s+f_y t)} ds dt \right|^2. \quad (S2)$$

Next, we convert the coordinate system from the Cartesian to polar by substituting  $f_x = \rho \cos \theta$ ,  $f_y = \rho \sin \theta$ ,  $s = r \cos \phi$ ,

$t = r \sin \phi$ . Accordingly, we further rewrite Eq. S2 as:

$$PSF(\rho, \theta, \lambda) = \left| \frac{1}{\lambda f} e^{i\pi\lambda f \rho^2} \int_0^\infty \int_0^{2\pi} P(r, \phi, \lambda) e^{\frac{ik}{2f}r^2} e^{-i2\pi(\rho \cos \theta r \cos \phi + \rho \sin \theta r \sin \phi)} r d\phi dr \right|^2 \quad (S3)$$

$$= \left| \frac{1}{\lambda f} e^{i\pi\lambda f \rho^2} \int_0^\infty \int_0^{2\pi} P(r, \phi, \lambda) e^{\frac{ik}{2f}r^2} e^{-i2\pi\rho r \cos(\theta-\phi)} r d\phi dr \right|^2.$$

by assuming the height profile of DA is rotational symmetric, we observe the polar version of the pupil function  $P(r, \phi, \lambda)$  is

independent with  $\phi$ . Then, Eq. S3 can be rewritten as:

$$\begin{aligned}
PSF(\rho, \lambda) &= \left| \frac{1}{\lambda f} e^{i\pi\lambda f \rho^2} \int_0^\infty \int_0^{2\pi} P(r, \lambda) e^{\frac{ik}{2f} r^2} \right. \\
&\quad \left. e^{-i2\pi\rho r \cos(\theta-\phi)} r d\phi dr \right|^2 \\
&= \left| \frac{1}{\lambda f} e^{i\pi\lambda f \rho^2} \int_0^\infty r P(r, \lambda) e^{\frac{ik}{2f} r^2} \right. \\
&\quad \left. \int_0^{2\pi} e^{-i2\pi\rho r \cos(\theta-\phi)} d\phi dr \right|^2 \quad (S4) \\
&= \left| \frac{1}{\lambda f} e^{i\pi\lambda f \rho^2} \int_0^\infty r P(r, \lambda) e^{\frac{ik}{2f} r^2} 2\pi J_0(2\pi\rho r) dr \right|^2 \\
&= \left| \frac{2\pi}{\lambda f} e^{i\frac{k}{2f} (\lambda f \rho)^2} \int_0^\infty r P(r, \lambda) e^{\frac{ik}{2f} r^2} J_0(2\pi\rho r) dr \right|^2,
\end{aligned}$$

where  $J_0$  is the 0<sup>th</sup> order of Bessel function of the first kind.

Since the DA is discretized as a number of concentric rings with a ring width of  $d$ , as illustrate in Fig. 2 in main text, the term  $P(r, \lambda) e^{\frac{ik}{2f} r^2}$  can be decomposed to many **circ**, written as:

$$\begin{aligned}
P(r, \lambda) e^{\frac{ik}{2f} r^2} &\approx P(r_1, \lambda) e^{\frac{ik}{2f} r_1^2} \text{circ} \left( \frac{r}{r_1} \right) + \\
&\quad \sum_{m=2}^\infty [P(r_m, \lambda) e^{\frac{ik}{2f} r_m^2} \text{circ} \left( \frac{r}{r_m} \right) - \\
&\quad P(r_m, \lambda) e^{\frac{ik}{2f} r_m^2} \text{circ} \left( \frac{r}{r_{m-1}} \right)], \quad (S5)
\end{aligned}$$

where  $r_m = md$  (for  $m = 1, 2, \dots$ ),  $d$  is the feature size of DA. The approximation of Eq. S5 is accuracy when the ring sampling  $d$  sufficiently fine to accurately approximate the additional phase term  $e^{\frac{ik}{2f} r^2}$  by  $e^{\frac{ik}{2f} r_m^2}$ . for example  $d \leq \frac{\lambda}{2NA}$  [1], where  $NA$  is the numerical number of the DA. The **circ** is the unit **circ** function, defined as:

$$\text{circ}(r) = \begin{cases} 1, & |r| \leq 1 \\ 0, & |r| > 1 \end{cases}. \quad (S6)$$

By substituting Eq. S5 into Eq. S4, we derive the notations  $G(r, \lambda)$ ,  $K(r, \lambda)$  to represent  $P(r, \lambda) e^{\frac{ik}{2f} r^2}$  and  $\frac{2\pi}{\lambda f} e^{i\frac{k}{2f} (\lambda f \rho)^2}$ , respectively, expressed as:

$$\begin{aligned}
PSF(\rho, \lambda) &= \left| K(r, \lambda) \left[ \int_0^\infty r G(r_1, \lambda) \text{circ} \left( \frac{r}{r_1} \right) J_0(2\pi\rho r) dr + \right. \right. \\
&\quad \left. \sum_{m=2}^\infty G(r_m, \lambda) \int_0^\infty r \left[ \text{circ} \left( \frac{r}{r_m} \right) - \right. \right. \\
&\quad \left. \left. \text{circ} \left( \frac{r}{r_{m-1}} \right) \right] J_0(2\pi\rho r) dr \right] \right|^2. \quad (S7)
\end{aligned}$$

Eq. S7 can be further expressed as:

$$\begin{aligned}
PSF(\rho, \lambda) &= \left| K(r, \lambda) \left[ G(r_1, \lambda) \int_0^{r_1} r J_0(2\pi\rho r) dr + \sum_{m=2}^\infty G(r_m, \lambda) \right. \right. \\
&\quad \left. \left. \left[ \int_0^{r_m} r J_0(2\pi\rho r) dr - \int_0^{r_{m-1}} r J_0(2\pi\rho r) dr \right] \right] \right|^2 \\
&= \left| K(r, \lambda) \left[ G(r_1, \lambda) \frac{1}{2\pi\rho} r_1 J_1(2\pi\rho r_1) + \frac{1}{2\pi\rho} \sum_{m=2}^\infty G(r_m, \lambda) \right. \right. \\
&\quad \left. \left. [r_m J_1(2\pi\rho r_m) - r_{m-1} J_1(2\pi\rho r_{m-1})] \right] \right|^2 \\
&= \left| \frac{2\pi}{\lambda f} e^{i\frac{k}{2f} (\lambda f \rho)^2} \sum_{m=1}^\infty P(r_m, \lambda) e^{\frac{ik}{2f} r_m^2} H(r_m, \rho) \right|^2, \quad (S8)
\end{aligned}$$

**Table S1.** Configuration of the base unit of the utilized Res-Unet. Specifically, “conv-n(a)-k(b)-d(c)” represents a convolution layer with “a” output channels, using a “b × b” kernel, and using a stride of “c”. Each “Leaky Relu” has the slope 0.2. Each “Upsampling” represents the nearest neighbor upsampling with a factor 2 followed by a convolution layer with a 3 × 3 kernel.

Layer	Convolution layer	Activation
<b>Input</b>		
0	conv-n32-k3-d1	Leaky Relu
1	conv-n32-k4-d2	Leaky Relu
2	conv-n64-k4-d2	Leaky Relu
3	conv-n128-k4-d2	Leaky Relu
4	conv-n256-k4-d2	Leaky Relu
5-1	conv-n512-k3-d1	Leaky Relu
5-2	conv-n512-k3-d1	Leaky Relu
<b>Upsampling &amp; Concat.</b>		
6	conv-n256-k3-d1	Leaky Relu
<b>Upsampling &amp; Concat.</b>		
7	conv-n128-k3-d1	Leaky Relu
<b>Upsampling &amp; Concat.</b>		
8	conv-n64-k3-d1	Leaky Relu
<b>Upsampling &amp; Concat.</b>		
9	conv-n32-k3-d1	Leaky Relu
10	conv-n3-k3-d1	Leaky Relu
<b>Adding with the input</b>		

where  $H(r_m, \rho)$  is expressed as:

$$H(r_m, \rho) = \begin{cases} \frac{1}{2\pi\rho} [r_m J_1(2\pi\rho r_m) - r_{m-1} J_1(2\pi\rho r_{m-1})], & m > 1 \\ \frac{1}{2\pi\rho} r_1 J_1(2\pi\rho r_1), & m = 1 \end{cases}. \quad (S9)$$

## 2. DETAILS OF IMAGE RECOVERY NEURAL NETWORK

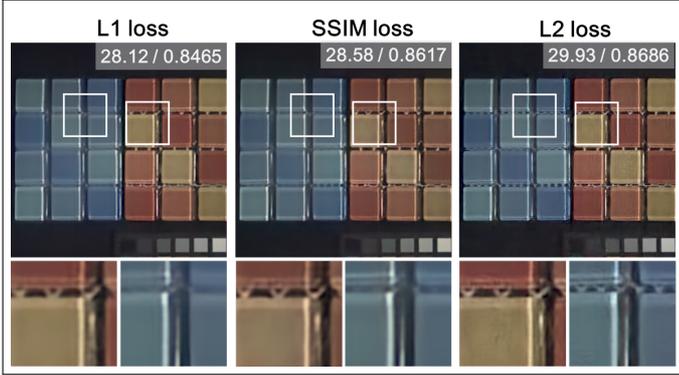
The configuration details of the base unit of the image recovery neural network are presented in Tab. S1.

## 3. ABLATION STUDY OF LOSS FUNCTIONS

We provide alternative results implemented with the  $\ell_1$  loss and SSIM loss. We use the same training set, parameters, and network configure except the loss function is changed to  $\ell_1$  loss and SSIM loss. We assess 15 test images using peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and spectral angular mapper (SAM) [2]. The Gaussian noise with  $\sigma=0.003$  is added to all sensor measurements. Table S2 summarizes the average PSNR, SSIM, SAM and Fig. S1 shows an example selected from the test set. We observe that the  $\ell_2$  loss leads to better results (with the average PSNR improvements of 0.50 dB).

**Table S2.** Quantitative evaluation of the averaged PSNR (dB), SSIM, and SAM over 15 test images resolved with the network trained with different loss functions.

Loss	$\ell_1$	SSIM	$\ell_2$
Recovery	32.53 / 0.886 / 0.07	32.59 / <b>0.896</b> / 0.07	<b>33.09</b> / 0.892 / <b>0.07</b>



**Fig. S1.** Assessment under different loss functions in simulation. We assess the performance of  $\ell_1$  loss, SSIM loss, and  $\ell_2$  loss under same training set, parameters, and network configuration. We show the recovery results of the Res-Net. The insert values indicate the PSNR (dB) and SSIM.

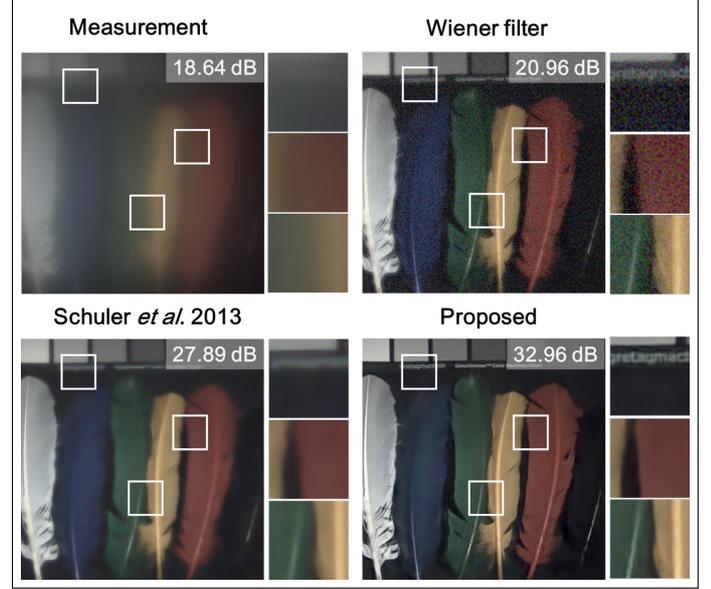
#### 4. ADDITIONAL SIMULATION RESULTS

**Comparison with alternative image recovery algorithms** We show the comparison among different image recovery methods. Table S3 summarizes the averaged PSNR and Fig. S2 visualizes several examples selected from the test set. We observe that the score of Wiener filtering is quite low since the recovered images are noisy (see Fig. S2), although the details and the color fidelity look good. After further applying a learning-based multi-layer perceptron (MLP) approach [3], the score shows a gain of 3.33 dB. However, this score is still significantly lower than that of images recovered by the proposed Res-Net (see Fig. S2). These results validate that the proposed Res-Net is capable of preserving high-fidelity scene details while suppressing noise in the meantime.

**Table S3.** Quantitative evaluation of averaged PSNR (dB) over 15 test images resolved using different recovery algorithms.

Algorithm	Wiener filter	Schuler et al. [3]	Res-Net
Recovery image	23.38	26.71	<b>33.09</b>

**Comparison under different noise levels** We provide additionally results for the performance assessment of the diffractive achromat (DA) designs with noise levels  $\sigma=0.006$  and  $\sigma=0.011$ . Table S4 summarizes the averaged PSNR and Fig. S3 visualizes an example selected from the test set. We observe that the proposed DA, in tandem with the Res-Net, obtains the best performance under noise levels of 0.003 and 0.006. Note that this performance improvement becomes less noticeable when the noise level is increased. Specifically, the PSNR improvement decreases from 1.3 dB to 0.16 dB, and the SSIM improvement decreases from



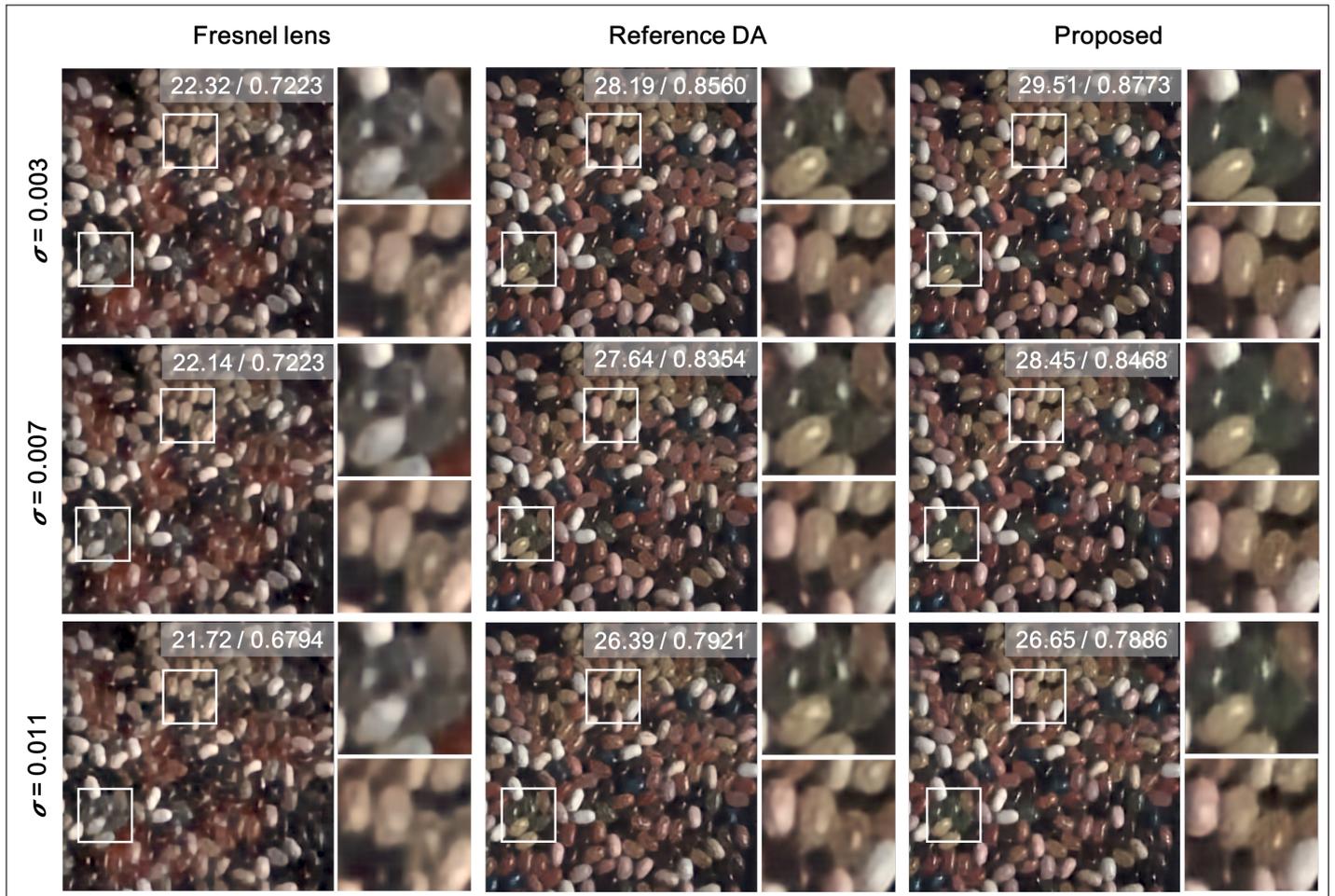
**Fig. S2.** Selected examples for performance evaluation of the image recovery methods in simulation. We compare against the recovery results of learned recovery (Res-Net network), a modified Wiener filter, and a modified Wiener filter combined with a learning-based multi-layer perceptron (MLP) approach [3]. The insert values indicate the PSNR (dB).

0.015 to N/A. That said, our current end-to-end design framework may only show superior results when deriving DAs for full-spectrum applications under low and median noise levels, which are very reasonable for state-of-the-art sensors.

**Table S4.** Quantitative evaluation of averaged PSNR (dB), SSIM, and SAM over 15 test images resolved with different lens designs and Res-Net recovery algorithms at different noise levels.

Noise Level	Fresnel Lens	Reference DA	Proposed DA
$\sigma=0.003$	25.78 / 0.804 / 0.12	31.79 / 0.877 / 0.08	<b>33.09 / 0.892 / 0.07</b>
$\sigma=0.006$	25.46 / 0.786 / 0.13	31.35 / 0.865 / 0.08	<b>32.30 / 0.874 / 0.07</b>
$\sigma=0.011$	24.78 / 0.734 / 0.15	30.38 / <b>0.840</b> / 0.08	<b>30.54 / 0.828 / 0.08</b>

**Comparison with a hyperspectral diffractive lens** We have compared with the DOE designed to better estimate (e.g., hyperspectral imaging) the spectral information as in [4]. We design the hyperspectral DOE using the method described in the relevant work [4] with a focal length of 50 mm and an aperture diameter of 8 mm. The design spectrum range is from 420 nm to 699 nm. We use the same training set and parameters to train the image recovery Res-Net for this hyperspectral DOE. We assess 15 test images using PSNR, SSIM, and SAM. The Gaussian noise with  $\sigma=0.003$  is added to all sensor measurements. Table S5 summarizes the average PSNR, SSIM, SAM and Fig. S4 shows an example selected from the test set. We observe that the proposed DA obtains better results (with the averaged PSNR and SAM improvements of 8.82 dB and 0.06, respectively). As shown in the right bottom of Fig. S4, the hyperspectral DOE is designed to distinguish the spectral PSFs as much as possible. Thus, the resulting blur is very sensitive to the aperture diameter.



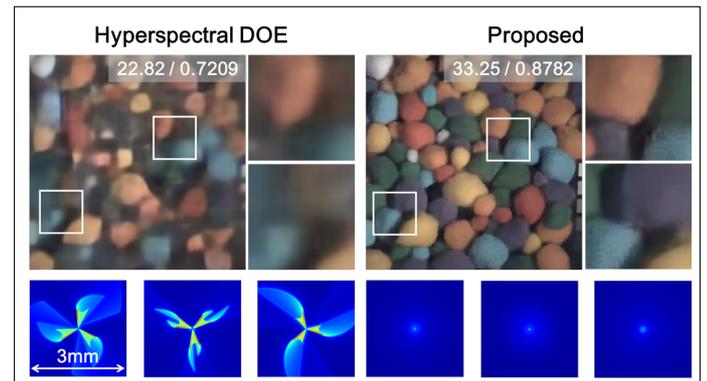
**Fig. S3.** Selected examples of the assessment of three lens designs with different noise levels in simulation. We show the recovery results of the Res-UNet. The inset values indicate the PSNR (dB) and SSIM.

In our case, the blur size of the PSFs goes up to approximately 3 mm, which causes image recovery algorithms failing to resolve a full-spectrum image. This means for a practical DOE with an aperture diameter of 8 mm, their design framework fails to obtain high-fidelity color images.

**Table S5.** Quantitative evaluation of averaged PSNR (dB), SSIM, and SAM over 15 test images resolved with different lens designs and different ResUNet recovery algorithms.

Type	Hyperspectral DOE	Proposed DA
Measurement	17.77 / 0.613 / 0.19	19.90 / 0.657 / 0.14
Recovery	24.27 / 0.714 / 0.13	33.09 / 0.892 / 0.07

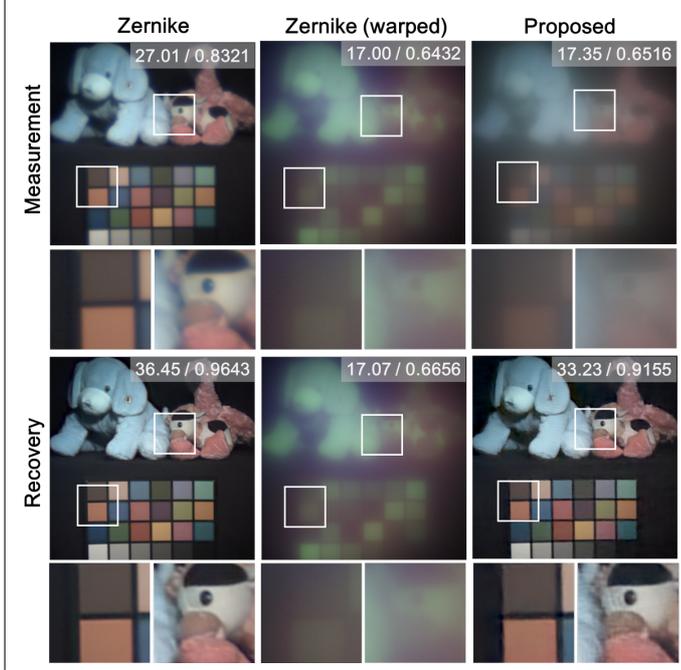
**Comparison with a DOE designed with alternative rotationally symmetric base representation** We choose the 16 circularly symmetric Zernike bases among the first 150 Zernike terms. For a fair comparison, we use the same training set, parameters, and network structure to learn the lens. We assess on 15 test images using the PSNR, SSIM, and SAM. The Gaussian noise with  $\sigma=0.003$  is added to all sensor measurements. Table S6 summarizes the averaged PSNR, SSIM, SAM, and Fig. S5 shows an example selected from the test set. Although the learned lens with Zernike



**Fig. S4.** Assessment between proposed lens designs and the hyperspectral DOE. The inset values indicate the PSNR (dB) and SSIM. We show the PSFs of two DOEs at the wavelengths of 447 nm, 555 nm, and 645 nm at the bottom, respectively. The PSFs shown here are gamma-corrected for visualization purpose.

basis shows the best result, it is a refractive lens because of its continual surface representation. That said, the maximum height of the lens is around  $284 \mu\text{m}$ , which is  $142\times$  as that of the

designed DOE. If we warp the learned Zernike lens to the same height of the proposed DA, as shown in the middle of Fig. S5, the warped Zernike lens becomes a Fresnel lens, further leading to poor performance.



**Fig. S5.** Selected examples of the assessment between the proposed DA and the DOE designed with alternative rotationally symmetric basis. The inset values indicate the PSNR (dB) and SSIM. Warping means to warp the learned Zernike lens to the same height of the proposed DA.

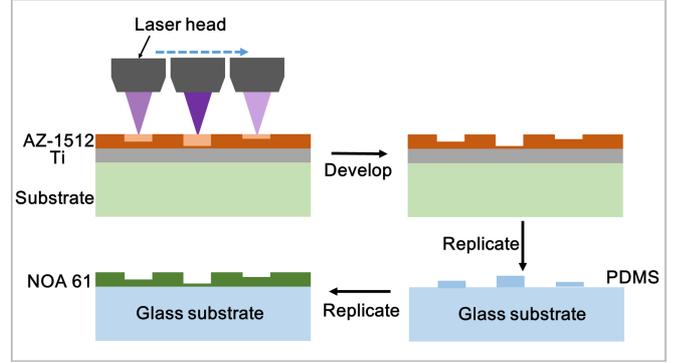
**Table S6.** Quantitative evaluation of averaged PSNR (dB), SSIM, and SAM, over 15 test images resolved with different lens designs and ResUnet recovery algorithms.

Type	Zernike	Zernike (warped)	Proposed DA
Measurement	27.89 / 0.803 / 0.08	19.58 / 0.647 / 0.17	19.90 / 0.657 / 0.14
Recovery	36.56 / 0.940 / 0.05	19.66 / 0.672 / 0.16	33.09 / 0.892 / 0.07

## 5. PROTOTYPING

**Fabrication details** As shown in Fig. S6, the designed DOEs are fabricated using the imprinting lithography technique. Specifically, the positive photoresist (AZ-1512, MicroChemicals) is spun on a titanium-coated glass substrate and is patterned by a direct-write optical grayscale lithography machine (MicroWriter ML3, Durham Magneto Optics). Then, the photoresist is developed with a base developer (MF-319, Microposit) and is used as a mold to replicate its pattern on polydimethylsiloxane (PDMS, SYLGARD 184, Dow). PDMS, a two-part polymer, is mixed at the standard ratio of 10:1 and degassed in the vacuum. The mixed PDMS is cured at room temperature for three days with the master mold to make the elastomeric mold. The circular aperture is fabricated with a light-blocking chromium-gold-chromium tri-layer (50 nm / 100 nm / 50 nm) on a 3 mm thick float glass substrate (30-773, Edmund Optics) through the lift-off process.

Then, a drop of optically clear UV-curable resin (NOA61, Norland Products) is put between the PDMS mold and the glass substrate and cured with a mercury vapor lamp after aligning the pattern with the circular aperture. Finally, the PDMS mold is gently peeled off to form a patterned clear resin layer on the glass substrate.



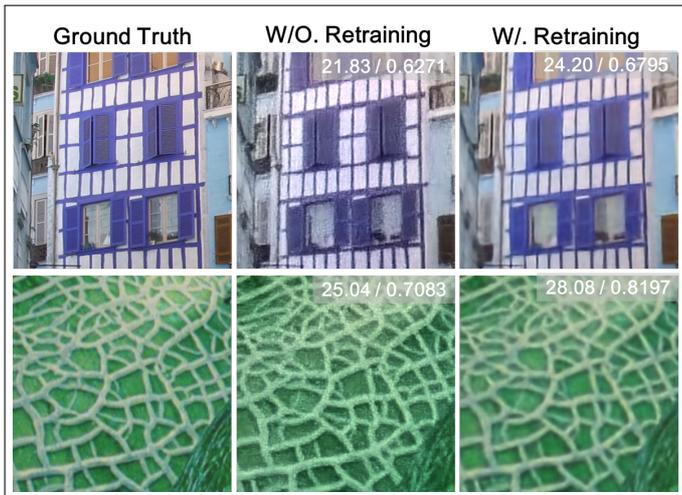
**Fig. S6.** Fabrication process of our proposed DA.

**Prototype camera and experiment setup** As shown in Fig. S7, the fabricated DA is mounted to a 1-inch stand tube and then attached to a Canon T5i DSLR camera body via a customized printed holder. We capture the PSF of the fabricated DA using a white LED light source and a 35  $\mu\text{m}$  pinhole attached in front. The distance between the light source and the front surface of the fabricated DA is 3 m, and we slightly adjust the gap between the DA and DSLR sensor to focus at the light source. To obtain a PSF with a high dynamical range, we capture PSFs with four integration settings, including 100 ms, 200 ms, 400 ms, and 800 ms, and then merge them. The measured PSF is used to refine the image recovery neural network.



**Fig. S7.** Photographs of the prototype lens and the experimental setup using a DSLR camera body.

**Image recovery network refining** In Fig. S8, we show the selected results with and without the refining of the image recovery network using the measured PSFs. It is clear that the refining step leads to an increase of performance. Here, we use a new RGB dataset [5] since we have replaced the spectral PSFs at 31 wavelengths with the RGB PSF. 20 images are used for the evaluation and the remaining for refining. Finally, we assess on 20 images with the Gaussian noise level of 0.005 and observe that both the PSNR and SSIM have indicated the increase from 27.88 dB and 0.774 to 31.03 dB and 0.793, respectively. The visual performance, regarding color performance and noise issue, has been significantly improved as well.



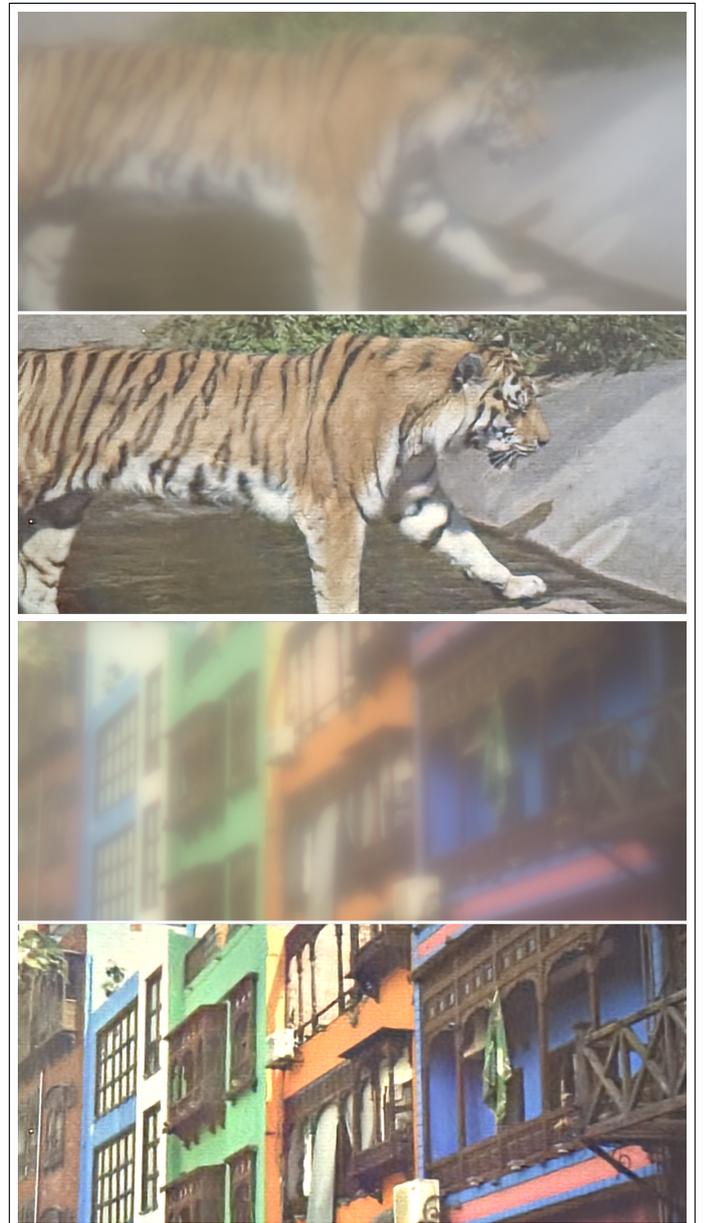
**Fig. S8.** Selected results with and with out the refining of the image recovery network using the measured PSFs. The inset values indicate the PSNR (dB) and SSIM.

## 6. ADDITIONAL CAPTURED RESULTS

To comprehensively verify the achromatic performance of our designed DA, we evaluate our designed DA on another machine vision sensor (Pointgrey Grasshopper3 USB3) that has  $1,900 \times 1,200$  pixels with the pixel pitch of  $5.9 \mu\text{m}$ . We reimplement the PSF calibration and refine the image recovery network, as mentioned in the main text. Figure S9 shows several results that captured in the real world where the images are displayed on a display monitor. The details and color are well resolved, indicating that our jointly designed diffractive lens suits well over different spectral response curves of sensors due to its robust achromatic behavior.

## REFERENCES

1. 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).
2. 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.* **283**, 192–201 (1993).
3. C. J. Schuler, H. Christopher Burger, S. Harmeling, and B. Scholkopf, "A machine learning approach for non-blind image deconvolution," in *Proceedings of CVPR, 2013 IEEE International Conference on*, (2013), pp. 1067–1074.
4. 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).
5. H. Jegou, M. Douze, and C. Schmid, "Hamming embedding and weak geometric consistency for large scale image search," in *Computer Vision – ECCV 2008*, (2008).



**Fig. S9.** Experimental results of the fabricated DA. For each pair, we show the degraded sensor measurement and the recovery result. The exposure time for these images are set 100 ms with gain 0. The processing time on an NVIDIA 1080Ti GPU is around 1 second.