# Kernelless Blind Inverse Imaging for Flat Meta-Optics Camera

Samuel Pinilla,*, Wenzhu Xing, † Seyyed Reza Miri Rostami, † Vladimir Katkovnik,†
Igor Shevkunov,† Johannes E. Fröch, ‡ Arka Majumdar, ‡Karen Egiazarian, †
* Science and Technology Facilities Council, Harwell, UK
†Tampere University, Tampere, Finland
‡ University of Washington, Seattle, WA, USA

***Abstract*—The race for color micro-cameras employing flat meta-optics instead of conventional refractive lenses has rapidly developed various end-to-end design frameworks. The meta-optics produce a specially designed spatial modulation of light wavefronts resulting in heavily blurred registered images. The optimal modulation is engineered to achieve advanced sharp imaging after computational data processing. The wavefront modulation and the image reconstruction are the fundamental micro-camera design problems with meta-optics. The popular convolution-based blurred image modeling (kernel-based) does not fit well with cameras with meta-optics. As a valuable alternative, we develop for image reconstruction the kernelless blind inverse imaging. This technique is based on a convolutional neural network. Its efficiency is demonstrated in the frame of the hardware-in-the-loop (HIL) joint optimization of meta-optics and image reconstruction software. The developed HIL setup allows us to overcome fundamental limitations of mismatch between theory-based and resultant experimental image formation problems of meta-optics. The resulting camera achieves high-quality full-color imaging for a $5$ mm aperture optics with a focal length of $5$ mm. We have observed a superior quality of the images captured by the developed hybrid meta-optical camera compared to the compound multi-lens optics of a commercial camera.**

## I. Introduction

The miniaturization of cameras, while maintaining high image quality, has become a major driving force in optics and photonics research [1]. A solution to this conundrum is *computational imaging*, where a digital backend augments the deficiencies of the optical components and improves the image quality [2]. A promising approach is to co-optimize the meta-optics, and the computational backend in an "end-to-end" design framework, where the hardware and software are equally considered, thus ensuring optimal system-level performance [3]–[6].

There are three potential advantages of meta-optics: 1) they can achieve significant size and weight reduction (thickness in micrometers) [7]; 2) performance beyond conventional systems is achievable in extended imaging modalities, such as extended-depth-of-field [8]–[10], and face recognition [11]; 3) nearly arbitrary sub-wavelength resolution manipulation of wavefields due to the immense progress in nanofabrication over the recent decades. Despite these promises, the image quality of captured images using large-aperture ($>$ 1 mm) meta-optics is plagued with strong chromatic and geometric aberrations [12], [13]. While complicated meta-atom engineering can help, fabricating large aperture meta-optics with small features remains challenging. Also, the design tools to create large aperture meta-optics are computationally expensive, which poses a severe challenge, as the current state-of-the-art "end-to-end" design approach [4], [14], [15] largely involves numerical modeling of the optics.

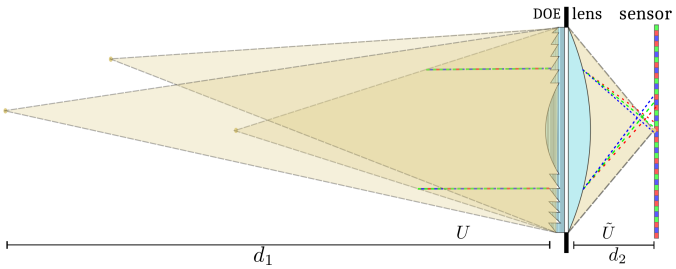


Fig. 1. Hybrid optics of imaging system. Light waves from different distances $d_1$ propagate on the aperture plane containing DOE to be designed. The DOE modulates the phase of the incident wavefront. The resulting wavefront propagates through the lens to the aperture-sensor, distance $d_2$.

In this work, we circumvent all these challenges by employing a hybrid refractive/meta-optics system (see Fig. 1) with a computational backend, following a recently proposed design methodology [2]. A hardware-in-the-loop strategy first optimizes the desired phase profile (see Fig. 2), where the DOE, implemented by a spatial light modulator, is configured and updated, while a refractive lens and a sensor remain fixed. This bypasses the expensive computational requirements to design the optics, directly accounting for sensor noise and precludes calibration errors. The optimized phase-mask is then implemented using meta-optics. Using hybrid optics over meta-optics-only systems improves the low light efficiency of meta-optics [16], aberrations, and imaging quality [17], [18].

Simple meta-atoms are employed to ensure fabricability, and the meta-atoms are designed only at $510$ nm, where the spatial light modulator operates. Although only a single wavelength is considered, we still achieve high-quality full-color imaging, as the meta-optics extend the depth-of-focus [19]. Thus, even using simple meta-atoms optimized for only one wavelength, we can realize high-quality broadband imaging thanks to hybrid optics and a computational backend. The resulting imaging systems have a focal length and aperture of $5$mm. To illustrate the potential of this camera, we compare it with a single lens-only configuration and the compound multi-lens optics of the Sony Alpha 1 III mirrorless commercial camera.

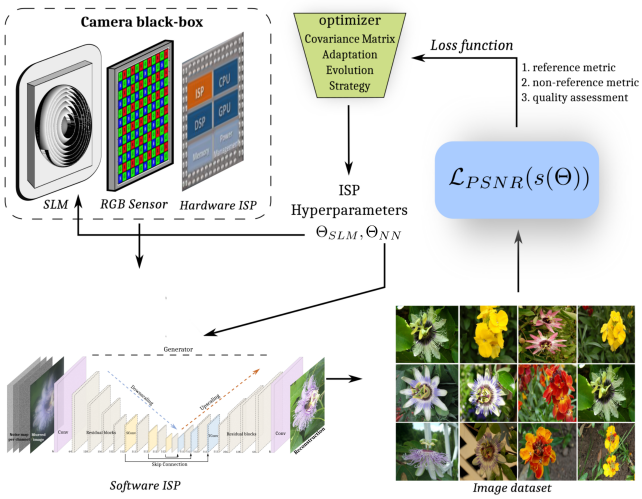


Fig. 2. Schematic of the hardware-in-the-loop methodology to co-optimize the encoding phase distribution and image reconstruction algorithm (Software ISP) in an end-to-end framework. The camera black-box model is composed of a spatial light modulator (SLM), which updates the encoding phase in optimization iterations, while sensor and hardware ISP are fixed. The output of this black-box forms the input for the image processing algorithm (Software ISP).

This comparison confirms the efficiency of the meta-hybrid optics in full-color imaging, whereas our designed hybrid meta-optics camera volume is approximately $108\times$ smaller than the mirrorless camera.

## II. End-to-end optimization of meta-optics and image reconstruction

In this section we describe a nonlinear black-box optimizer (summarized in Algorithm 1) to design the hybrid optics.

### A. Computational design and imaging

Algorithm 1 follows an alternating methodology: fixing hyperparameter $\Theta_{SLM}$, solving $\Theta_{NN}$ for inverse imaging, updating $\Theta_{SLM}$ and so forth. We implement this methodology because it is faster than training the CNN-based inverse imaging per SLM-pattern iterations. This alternating process starts by randomly selecting an SLM-pattern $\Theta^{(0)}_{SLM}$ and a pre-trained inverse imaging CNN for a wide range of hyperparameter $\Theta_{SLM}$. After this, algorithm acquires a set of blurred images at distances from sensor $0.5, 1.0, 1.8$ m using the ISP of the optical system, $s(\Theta^{(r)}_{SLM})$, which is then passed to the downstream reconstruction (deblurring using trained CNN) module. The output of the task module is evaluated by domains-specific evaluation metric which in this case is the peak-signal-to-noise-ratio (PSNR), $\mathcal{L}_{HIL}(s(\Theta^{(r)}_{SLM}))$. Then, using the $0th$-order stochastic evolutionary search method CMA-ES [20] (documentation in python of the CMA-ES optimizer in https://pypi.org/project/cma/), Algorithm 1 updates $\Theta_{SLM}$ taking advantage of the tested SLM-patterns during the $R$ iterations. Once $\Theta_{SLM}$ is updated, Algorithm 1 refines the CNN-based inverse imaging by training it for the best $\Theta^{(r)}_{SLM}$.

Performing the previous alternating process $N_{iter}$ times algorithm returns the updated $\Theta_{SLM}$, and $\Theta_{NN}$. The structure

**Algorithm 1** HIL design and inverse imaging

**input**: $\lambda, N_{iter}$, and $\Theta^{(0)}_{SLM}$.
**initialize CMA-ES**: $\Theta_{SLM} \leftarrow \Theta^{(0)}_{SLM}, t \leftarrow 1$
Train initial inverse imaging CNN for $0.5, 1.0, 1.8$m, $\Theta_{NN} \leftarrow \Theta^{(0)}_{NN}$ for a wide range of hyperparameters $\Theta_{SLM}$
**while** $t \leq N_{iter}$ **do**
  **for** $r = 1$ to $R$ **do**
    $\Theta^{(r)}_{SLM} \leftarrow$ randomly draw from Gaussian at $\Theta_{SLM}$
    Add random noise to $\Theta^{(r)}_{SLM}$
    $s\left(\Theta^{(r)}_{SLM}\right) \leftarrow$ get blurred data at $0.5, 1.0, 1.8$m of images $\boldsymbol{I}_1, \ldots, \boldsymbol{I}_{\mathcal{J}}$ and use $\Theta^{(t)}_{NN}$-CNN to estimate them
    $\mathcal{L}_{HIL}\left(s\left(\Theta^{(r)}_{SLM}\right)\right) \leftarrow$ compute average PSNR among the estimated $\mathcal{J}$-images at each distance $0.5, 1.0, 1.8$m
    $\Theta_{SLM} \leftarrow$ update CMA-ES
  **end for**
  $t \leftarrow t + 1$
  $\Theta^{(t)}_{NN} \leftarrow$ train inverse imaging CNN for best SLM-pattern among $\{\Theta^{(1)}_{SLM}, \ldots, \Theta^{(\lambda)}_{SLM}\}$
**end while**
**return:** $\Theta_{SLM}, \Theta_{NN}$

of CNN developed for inverse imaging (optimization on $\Theta_{NN}$) is shown in Figure 2. Algorithm 1 can be also initiated by solutions obtained according to the model-based approach from [21]. The number of 'global' iterations of this algorithm for $\Theta_{NN}$ is $N_{iter}$ and $R$ is the number of 'local' iterations for $\Theta_{SLM}$, where the time needed to process physical observations is about 1.01 seconds. This processing time involves the deconvolution of the blur data is collected at the sensor, and the decision of a new SLM-pattern. In this work we fixed $N_{iter} = 3$ and $R = 500$. In the following sections, more details per each stage are described.

An appropriate loss function is required to optimize our inverse imaging to provide the desired output. Thus, we use a weighted combination of PSNR between estimated and ground truth images, $\mathcal{L}_{PSNR}$, and perceptual losses given below by:

**Perceptual loss:** To measure the semantic difference between the estimated output and the ground truth, we use a pretrained VGG-16 [22] model for our perceptual loss [23]. This is motivated by the fact that in [23] has been reported that VGG promotes sharp details such as edges, and rapid colour changes in an image making it suitable to design optics. To evaluate this metric we have to extract feature maps between the second convolution (after activation) and second max pool layers $\varphi_{22}$, and between the third convolution (after activation) and the fourth max pool layers $\varphi_{43}$. Then, the loss $\mathcal{L}_{Percep}$ is the averaged PSNR between the outputs of these two activation functions for both estimated and ground truth images. This means we are using second, third and fourth layers of the VGG which returns feature maps [22] to estimate the perceptual distance between the estimated and original

image while designing the optics. For more details about this semantic metric please refer to [22].

**Adversarial loss:** Adversarial loss [24] was added to further bring the distribution of the reconstructed output close to those of the real images. Given the swish activation function [25] as our discriminator $D$, this loss is given as $\mathcal{L}_{Adv} = -\log(D(I_{est}))$ where $I_{est}$ models the estimated image.

Our total loss for the proposed CNN inverse imaging while training is a weighted combination of the three losses and is given as, $\mathcal{L}_{CNN} = \sigma_1\mathcal{L}_{PSNR} + \sigma_2\mathcal{L}_{Percep} + \sigma_3\mathcal{L}_{Adv}$, where, $\sigma_1, \sigma_2$ and $\sigma_3$ are empirical weights assigned to each loss. In this work these constant are fixed as $\sigma_1 = 1.0, \sigma_2 = 0.6$, and $\sigma_3 = 0.1$. Lastly, the parameters of this networks to be optimized are summarized in $\Theta_{NN}$.

*Hardware-in-the-loop Optimization of Hybrid Camera*

The HIL optimization is formalized using two models, one for the black-box camera and the second for the image reconstruction (inverse imaging) software. The observations registered by the sensor are defined as

$$y_i = \mathcal{T}_{hard}(\Theta_{SLM},\ x_i), \tag{1}$$

where $x_i$ is a true image, $y_i$ represents the blurred image collected at the sensor, and $\Theta_{SLM}$ is the phase profile of the flat optics at the SLM. The operator $\mathcal{T}_{hard}$ in Eq. (1) includes image formation due to light-beam propagation through optical components of the system and signal transformations by other hardware elements in the loop. The mathematical model of this operator is unknown, and $y_i$ corresponding to $x_i$ can be found experimentally only. The varying phase profile of SLM $\Theta_{SLM}$ is subject to optimization.

On the other hand, the inverse imaging software operator $\mathcal{T}_{NN}$ , parametrized by $\Theta_{NN}$, reconstructs an estimate of $x_i$ from $y_i$ . This process is modeled as

$$\widehat{x}_i = \mathcal{T}_{NN}(\Theta_{NN},\ y_i). \tag{2}$$

The design of hybrid camera requires to find the encoding phase modulation for DOE parametrized by $\Theta_{SLM}$ and the imaging software operator $\mathcal{T}_{NN}$ with parameter $\Theta_{NN}$.

To design the flat optics and inverse imaging software jointly, we use an end-to-end approach employing neural networks. To this end, the set of training images $\{x_i\}_{i=1}^N$ is given, the operator $\mathcal{T}_{NN}$ in Eq. (2) is defined by the structure of DRUNet [26], and $\Theta_{NN}$ are parameters of this network. This network can handle various noise levels for an RGB image, per channel, via a single model. The backbone of DRUNet is U-Net. Therefore, the goal of our end-to-end optimization is formulated as follows

$$(\widehat{\Theta}_{SLM}, \widehat{\Theta}_{NN}) \in \underset{\Theta_{SLM},\Theta_{NN}}{\arg\min} \sum_{i=1}^{N} \mathcal{L}(\widehat{x}_i, x_i), \tag{3}$$

where $\mathcal{L}$ is a loss function composed from peak signal-to-noise ratio (PSNR) values for accuracy evaluation and perceptual metrics. Remark that combining Eqs.1 and 2, we have that $\widehat{x}_i = \mathcal{T}_{NN}(\Theta_{NN},\ \mathcal{T}_{hard}(\Theta_{SLM},\ x_i))$ in Eq. (3). Once the hybrid camera is designed, the attained solution $(\widehat{\Theta}_{SLM}, \widehat{\Theta}_{NN})$ enables image reconstruction given as

$$\widehat{x}_i = \mathcal{T}_{NN}(\widehat{\Theta}_{NN}\ y_i). \tag{4}$$

Observe that the optics is fixed in Eq. (4), and it defines the inverse imaging for any $y_i$.

This image reconstruction algorithm approximates the inverse imaging operator for image formation by $\mathcal{T}_{hard}$. The algorithm can be classified as a blind inverse imaging algorithm as not using convolutional kernels invariant or varying. This method is kernelless which is cardinally different with respect to the conventional image inverse imaging algorithms using various models of convolutional kernels and point spread functions (PSFs) [1], [16], [27]–[29].

## III. Results

### A. Design and implementation

The optical system have been designed for achromatic extended depth of field (EDoF) imaging. To assess its performance numerically and visually we captured images in two scenarios: (1) *three monitor imaging* with three fixed distances $(0.5, 1.0, 1.8$ m) between the imaging monitors and sensor and (2) *real-life scene* with arbitrary locations of various colored objects at several distances relative to the camera (Fig.5). The three monitor scenario was used for the design: joint optimization of phase characteristics of meta-optics and image reconstruction CNN. The design of the phase modulation is composed from two successive problems. The first one is optimization of the SLM parameters in the frame of the developed HIL methodology. The optimized phase distribution is then directly translated to a meta-optic using simple square pillars in SiN (700 nm height, 350 nm period) on a quartz substrate (500 um thickness), which does not add a substantial thickness to the refractive lens (4.5 mm thickness). Fig. 3 provides a characterization of the general view and phase properties of the manufactured meta-lens, the latter in comparison with the optimized phase profile of SLM.

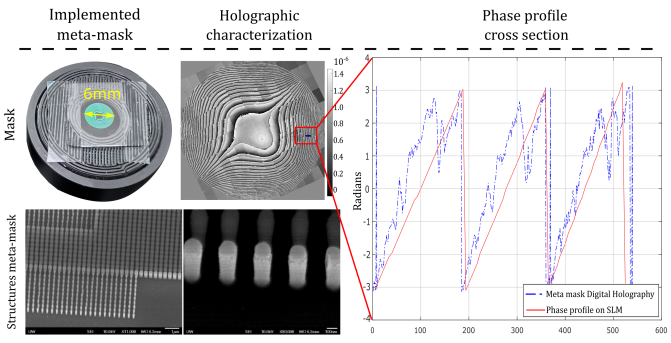


Fig. 3. Properties of the fabricated meta-mask. The bottom row show scanning electron micrographs highlighting the structural integrity across the device. The holographic characterization of the meta-mask phase-delay indicates a close match between design (red) and measurement (blue).

Importantly, although we consider the phase response for only a single wavelength (511 nm), the resultant meta-optic

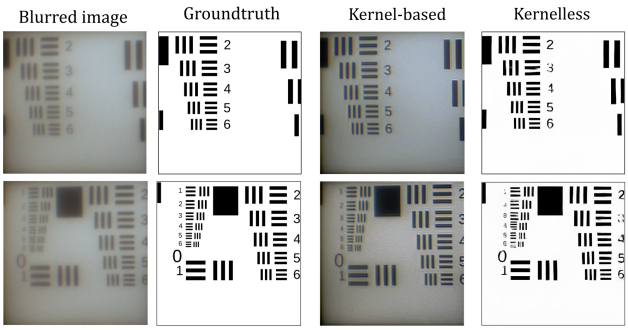


Fig. 4. In-focus monitor imaging. The first and second columns show the blurred (registered by sensor) and ground-truth images. The distance from object to sensor (focal distance) $d_1 = 1.0$ m. Columns 2 and 3 show the kernelless imaging with the proposed hybrid system and the kernel-based inverse imaging left and right, respectively.

still yields an improvement over the entire visible range as validated in full-color imaging experiments using a monitor setup (Fig.4), as well as under ambient light (Fig.5). In comparison to a single refractive lens-only configuration we achieve better image quality, particularly for off-focus depths. Even compared with a commercial compound lens we demonstrate on-par or better performance. An additional improvement of the designed hybrid meta-optics over the commercial compound lens, and the lens-only setups is the extended depth-of-field (see Fig.5). Essentially, by combining a refractive and meta-optic in one hybrid system we circumvent otherwise stringent limitations that otherwise restrict the performance of either system alone. Specifically, no meta-optic has been demonstrated up to date, with an equivalent combination of image quality, large aperture size, low $F$-number, and depth-of-field.

### *B. Imaging results*

The in-focus monitor imaging configuration results are illustrated in Fig.4. The first and second columns show the blurred and ground-truth images at a distance of $d_1 = 1.0$ m. Columns 3 and 4 show the imaging results obtained with kernel-based and kerneless inverse methodologies for the hybrid system. These reconstructions are obtained from the modulated, heavily blurred images captured by the sensor in column 1. In column 3, we show the images reconstructed from these blurred observations using the conventional type kernel-based Wiener inverse imaging. Here we use the algorithm presented in [21]. The comparison is clearly in favor of the developed kerneless technique.

In the real-life scene, we further compare the hybrid optic with a compound multi-lens camera, namely a Sony Alpha 1 III sensor with a Sony SEL85F18 lens (at the aperture limit values of $F/1.8$ and $F/22$). For the fair comparison, the focal distance of the multi-lens camera is fixed to 1 m as it is for the designed hybrid. The columns 1 and 2 show the images given by the camera. In the third column, one can see the corresponding images obtained by the hybrid. Specific image details, for objects at out-of-focus distances of 0.6 m and

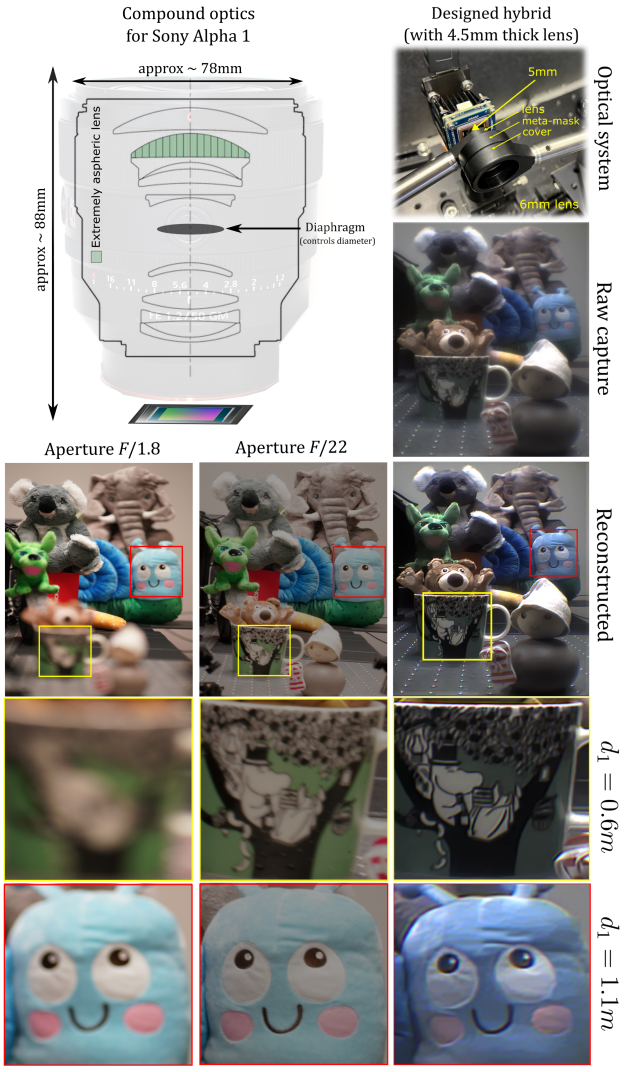


Fig. 5. Real-life scene composed of multiple objects/toys of various colors arranged at different depths relative to the sensor. The top left shows a schematic of the compound lens, whereas top right images show implementation of the hybrid system. The second row (right columns) show the raw capture of the hybrid system. The reconstructed images are presented in row 3 with enlarged fragments in rows 4 and 5, corresponding to off-focus distances of $d_1 = 0.6, 1.1$ m, respectively. The results from compound-lens camera are obtained at the limit aperture values of $F/1.8$ and $F/22$.

1.1 m, are enlarged in rows 4 and 5. The images obtained with a hybrid optic exhibit sharper images with a better resolution of details at both depths. Further, in comparison with the compound camera and aperture of $F/1.8$, the hybrid shows clearly better imaging quality across a larger depth range. Even for the smaller aperture value of $F/22$, the images obtained with the hybrid exhibit sharper details for out-of-focus depths at 0.6 m.

## IV. Conclusion

A full-color miniature computational camera composed of a refractive lens and large aperture meta-optic is demonstrated. The hybrid solution has an a low $F$-number of 1.0. A crucial

aspect in the development is the design of the meta-optic, where we employed a hardware-in-the-loop methodology to optimize the phase distribution and the computational backend in an end-to-end design framework. Specifically, the wavefront modulation and the image reconstruction are the fundamental micro-camera design problems with meta-optics. The popular convolution-based blurred image modeling (kernel-based) does not fit well with cameras with meta-optics. As a valuable alternative, we developed for image reconstruction the kernelless blind inverse imaging. This technique is based on a convolutional neural network. Its efficiency was demonstrated in the frame of the hardware-in-the-loop joint optimization of meta-optics and image reconstruction software. The developed HIL setup allows us to overcome fundamental limitations of mismatch between theory-based and resultant experimental image formation problems of meta-optics. We envision that the developed design concept, where a hardware-in-the-loop methodology is married with an kernelless computational backend in an end-to-end design framework, will be extended in future works towards diverse tasks, ranging from hyperspectral imaging, to classification or object detection tasks.

## ACKNOWLEDGMENT

This work is supported by the CIWIL project funded by Jane and Aatos Erkko Foundation, Finland, and by the Academy of Finland (project no. 336357, PROFI 6 - TAU Imaging Research Platform). Part of this work was conducted at the Washington Nanofabrication Facility/Molecular Analysis Facility, a National Nanotechnology Coordinated Infrastructure (NNCI) site at the University of Washington, with partial support from the National Science Foundation via Awards NNCI-1542101 and NNCI-2025489.

## REFERENCES

[1] E. Tseng, S. Colburn, J. Whitehead, L. Huang, S.-H. Baek, A. Majumdar, and F. Heide, "Neural nano-optics for high-quality thin lens imaging," *arXiv preprint arXiv:2102.11579*, 2021.

[2] S. Pinilla, S. R. M. Rostami, I. Shevkunov, V. Katkovnik, and K. Egiazarian, "Hybrid diffractive optics design via hardware-in-the-loop methodology for achromatic extended-depth-of-field imaging," *Optics Express*, vol. 30, no. 18, pp. 32 633–32 649, 2022.

[3] 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, 2021.

[4] S.-H. Baek, H. Ikoma, D. S. Jeon, Y. Li, W. Heidrich, G. Wetzstein, and M. H. Kim, "End-to-end hyperspectral-depth imaging with learned diffractive optics," *arXiv preprint arXiv:2009.00463*, 2020.

[5] 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 Transactions on Graphics (TOG)*, vol. 37, no. 4, pp. 1–13, 2018.

[6] A. Mosleh, A. Sharma, E. Onzon, F. Mannan, N. Robidoux, and F. Heide, "Hardware-in-the-loop end-to-end optimization of camera image processing pipelines," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020.

[7] J. E. Whitehead, A. Zhan, S. Colburn, L. Huang, and A. Majumdar, "Fast extended depth of focus meta-optics for varifocal functionality," *Photonics Research*, vol. 10, no. 3, pp. 828–833, 2022.

[8] S. MiriRostami, V. Y. Katkovnik, and K. O. Eguiazarian, "Extended DoF and achromatic inverse imaging for lens and lensless MPM camera based on Wiener filtering of defocused OTFs," *Optical Engineering*, vol. 60, no. 5, pp. 1 – 14, 2021. [Online]. Available: https://doi.org/10.1117/1.OE.60.5.051204

[9] Q. Fan, W. Xu, X. Hu, W. Zhu, T. Yue, C. Zhang, F. Yan, L. Chen, H. J. Lezec, Y. Lu *et al.*, "Trilobite-inspired neural nanophotonic light-field camera with extreme depth-of-field," *Nature communications*, vol. 13, no. 1, pp. 1–10, 2022.

[10] O. Lévêque, C. Kulcsár, A. Lee, H. Sauer, A. Aleksanyan, P. Bon, L. Cognet, and F. Goudail, "Co-designed annular binary phase masks for depth-of-field extension in single-molecule localization microscopy," *Optics Express*, vol. 28, no. 22, pp. 32 426–32 446, 2020.

[11] J. Tan, L. Niu, J. K. Adams, V. Boominathan, J. T. Robinson, R. G. Baraniuk, and A. Veeraraghavan, "Face detection and verification using lensless cameras," *IEEE Transactions on Computational Imaging*, vol. 5, no. 2, pp. 180–194, 2019.

[12] F. Presutti and F. Monticone, "Focusing on bandwidth: achromatic metalens limits," *Optica*, vol. 7, no. 6, pp. 624–631, 2020.

[13] L. Huang, S. Colburn, A. Zhan, and A. Majumdar, "Full-color metaoptical imaging in visible light," *Advanced Photonics Research*, vol. 3, no. 5, p. 2100265, 2022.

[14] Q. Sun, C. Wang, F. Qiang, D. Xiong, and H. Wolfgang, "End-to-end complex lens design with differentiable ray tracing," *ACM Trans. Graph*, vol. 40, no. 4, 2021.

[15] Y. Liu, C. Zhang, T. Kou, Y. Li, and J. Shen, "End-to-end computational optics with a singlet lens for large depth-of-field imaging," *Optics Express*, vol. 29, no. 18, pp. 28 530–28 548, 2021.

[16] 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 Transactions on Graphics*, 2019.

[17] M. S. Asif, A. Ayremlou, A. Sankaranarayanan, A. Veeraraghavan, and R. G. Baraniuk, "Flatcam: Thin, lensless cameras using coded aperture and computation," *IEEE Transactions on Computational Imaging*, vol. 3, no. 3, pp. 384–397, 2017.

[18] K. Monakhova, K. Yanny, N. Aggarwal, and L. Waller, "Spectral diffusercam: Lensless snapshot hyperspectral imaging with a spectral filter array," *Optica*, vol. 7, no. 10, pp. 1298–1307, 2020.

[19] S. Colburn, A. Zhan, and A. Majumdar, "Metasurface optics for full-color computational imaging," *Science Advances*, vol. 4, no. 2, p. eaar2114, 2018. [Online]. Available: https://www.science.org/doi/abs/10.1126/sciadv.aar2114

[20] N. Hansen and A. Ostermeier, "Adapting arbitrary normal mutation distributions in evolution strategies: The covariance matrix adaptation," in *Proceedings of IEEE international conference on evolutionary computation*. IEEE, 1996, pp. 312–317.

[21] S. R. M. Rostami, S. Pinilla, I. Shevkunov, V. Katkovnik, and K. Egiazarian, "Power-balanced hybrid optics boosted design for achromatic extendeddepth-of-field imaging via optimized mixed OTF," *Appl. Opt.*, vol. 60, no. 30, pp. 9365–9378, Oct 2021. [Online]. Available: http://www.osapublishing.org/ao/abstract.cfm?URI=ao-60-30-9365

[22] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[23] S. S. Khan, V. Sundar, V. Boominathan, A. Veeraraghavan, and K. Mitra, "Flatnet: Towards photorealistic scene reconstruction from lensless measurements," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

[24] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in neural information processing systems*, vol. 27, 2014.

[25] P. Ramachandran, B. Zoph, and Q. V. Le, "Searching for activation functions," *arXiv preprint arXiv:1710.05941*, 2017.

[26] K. Zhang, Y. Li, W. Zuo, L. Zhang, L. Van Gool, and R. Timofte, "Plug-and-play image restoration with deep denoiser prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[27] X. Dun, H. Ikoma, G. Wetzstein, Z. Wang, X. Cheng, and Y. Peng, "Learned rotationally symmetric diffractive achromat for full-spectrum computational imaging," *Optica*, vol. 7, no. 8, pp. 913–922, 2020.

[28] H. Arguello, S. Pinilla, Y. Peng, H. Ikoma, J. Bacca, and G. Wetzstein, "Shift-variant color-coded diffractive spectral imaging system," *Optica*, vol. 8, no. 11, pp. 1424–1434, 2021.

[29] H. Hu, H. Zhou, Z. Xu, Q. Li, H. Feng, Y. Chen, T. Jiang, and W. Xu, "Practical snapshot hyperspectral imaging with doe," *Optics and Lasers in Engineering*, vol. 156, p. 107098, 2022.