Hybrid Diffractive Optics Design via Hardware-in-the-Loop Methodology for Achromatic Extended-Depth-of-Field Imaging

SAMUEL PINILLA, University of Manchester at Harwell Science and Innovation campus, UK
SEYYED REZA MIRI ROSTAMI, Tampere University, Finland
IGOR SHEVKUNOV, Tampere University, Finland
VLADIMIR KATKOVNIK, Tampere University, Finland
KAREN EGUIAZARIAN, Tampere University, Finland

End-to-end optimization of diffractive optical elements (DOEs) profile through a digital differentiable model combined with computational imaging have gained an increasing attention in emerging applications due to the compactness of resultant physical setups. Despite recent works have shown the potential of this methodology to design optics, its performance in physical setups is still limited and affected by manufacturing artifacts of DOE, mismatch between simulated and resultant experimental point spread functions, and calibration errors. Additionally, the computational burden of the digital differentiable model to effectively design the DOE is increasing, thus limiting the size of the DOE that can be designed. To overcome the above mentioned limitations, the broadband imaging system with phase-only spatial light modulator (SLM) as an encoded diffractive optical element is proposed and developed in this paper. The SLM is exploited as a pixel-wise programmable device for design of a light-modulation phase pattern. A co-design of the SLM phase pattern and image reconstruction algorithm is produced following the end-to-end strategy, using for optimization a convolutional neural network equipped with quantitative and qualitative loss functions. The optics of the imaging system is hybrid consisting of SLM as DOE and refractive lens. SLM phase-pattern is optimized by applying the Hardware-in-the-loop technique, which helps to eliminate the mismatch between numerical modeling and physical reality of image formation as light propagation is not numerically modeled but is physically done. In our experiments, the hybrid optics is implemented by the optical projection of the SLM phase-pattern on a lens plane for a depth range 0.4-1.9 m (Figure 1). Multiple numerical and physical experiments confirm high-quality imaging of the system in the achromatic extended depth-of-field scenario. Comparison with compound multi-lens optics such as Sony A7 III and iPhone Xs Max cameras show that the proposed system is advanced in all-in-focus sharp imaging.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ACM Reference Format:
Samuel Pinilla, Seyed Reza Miri Rostami, Igor Shevkunov, Vladimir Katkovnik, and Karen Eguiazarian. 2022. Hybrid Diffractive Optics Design via Hardware-in-the-Loop Methodology for Achromatic Extended-Depth-of-Field Imaging. ACM Trans. Graph. 1, 1 (March 2022), 9 pages. https://doi.org/10.1145/mnnnnn.mnnnnn

Additional Key Words and Phrases: Hardware-in-the-loop design; Hybrid diffractive optics; Achromatic extended-depth-of-field.
1 INTRODUCTION

Computational imaging with encoded diffractive optical elements (DOEs) (e.g. binary-, multi-level phase elements) and meta-optical elements (MOEs) is a multidisciplinary research field in the intersection of optics, mathematics and digital image processing [Arguello et al. 2021; Chang et al. 2018; Colburn and Majumdar 2021; Dun et al. 2020; Sitzmann et al. 2018]. It is based on a combination of optical encoding (at the optics layer) and algorithmic decoding (at the image processing layer). Contrary to the traditional optical systems with refractive lenses, the hardware decoding (optical focusing) is replaced by the software computational decoding [Antipa et al. 2018; Asif et al. 2017; Boominnathan et al. 2016; Mosleh et al. 2020]. In DOE, this phase coding is induced due to the length of the ray path inside a DOE material. In MOE, the phase delay is induced via the response of nanostructures (so-called nanoantennas) built on the surface of the substrate material [Engelberg and Levy 2020]. Both DOEs and MOEs are used as the wavefront coding instruments [Chen et al. 2020b]. Despite the fundamental difference in nature between DOEs and MOEs, some authors treat MOEs as a special class of DOEs, and we also use DOE as a notation for both classes of these elements.

The potential of DOEs as optical elements is three-fold: (1) Accuracy and quality of imaging can be high at least on the level of the compound conventional refractive optics, while DOEs are compact (thickness in micrometers), light, and cheap. (2) The systems with DOEs have the potential to allow solutions beyond the ability of the conventional systems, in particular, for some pattern recognition problems, hyperspectral imaging, and even in such conventional types of problems as extended depth of field (EDoF) [MiriRostami et al. 2021], extended field-of-view, and achromatic broadband imaging. Typical examples are flat hyperspectral cameras [Monakhova et al. 2020] and flat face recognition devices [Tan et al. 2019]. (3) Nearly arbitrary manipulations of wavefields are possible due to amazing progress in nanotechnology. While nanotechnology allows obtaining DOEs with nearly arbitrary phase-patterns, the desirable optical characteristics as a rule are a priori unknown and an ordinary intuition usually cannot help, what makes the design of encoded DOEs quite problematic. Images registered by sensors in systems with encoded DOEs are blurred (assumed to be convolution between the knowledge of the gradient to optimize the optics. The phase-pattern as a linear combination of the first fourteen Zernike polynomials where coefficients are summarized in \( \Theta \)) is a zero-order algorithm, following a parameterized model of SLM [37]. The crucial advantage of CMA-ES for the HIL setup is that it does not require the knowledge of the gradient to optimize the optics. The phase-pattern of SLM is designed as a piece-wise invariant Multilevel Phase-Pattern (MPP) defined for the design wavelength. The software ISP is designed using CNNs as a natural tool for optimization of image reconstruction algorithms with respect to the hyperparameter \( \Theta_{\text{soft}} \) using Adam algorithm [Kingma and Ba 2014].

We assume that the hardware ISP, in particular including de-aliasing, is fixed. Other components of the hardware (sensor and hardware ISP: low-level image processing present in most digital cameras) also can be a subject of optimization provided that they are variable. Together with SLM and sensor, it outlines a camera as a ‘black-box’ (given as an input-output system of an unknown mathematical model) embedded in algorithmic optimization. Considering this paper, we propose an end-to-end design of the DOE through a ‘Hardware-In-the Loop (HIL)’ imaging setup, for achromatic EDoF, using a programmable phase Spatial Light Modulator (SLM) for implementation of encoded DOEs as illustrated in Figures 1 and 2.
the advantages of our HIL setup we have designed a DOE of 9.2mm diameter with a pixel resolution of 3.74μm and 10mm focal length. To the best of our knowledge these DOE parameters have not been achieved using state-of-the-art digital differentiable models due to limitations on computational costs (e.g. [Dun et al. 2020]).

In experiments for all-in-focus imaging, we compare the designed system with conventional compound multi-lens cameras such as iPhone Xs Max and Sony A7 III with the objective lens configured with 85mm focal length and F22. For the Sony camera, we chose the largest F-number (corresponding to the smallest aperture size) to result in the deepest DoF. Our system with hybrid optics designed in HIL strategy provides performance that is competitive in terms of imaging quality and advanced achromatic sharp imaging.

The contribution of this work can be itemized as follows:

- Broadband diffractive imaging with programmable SLM phase pattern projected to the lens plane with online physical tuning of DOE (for the first time);
- The framework for HIL co-design methodology of multilevel SLM phase-pattern (MPP) and inverse imaging for achromatic EDoF;
- Advanced performance in terms of EDoF for the developed imaging system in comparison with compound optics of SONY A7 III (85mm/F22.0) and iPhone Xs Max;

Scope. We believe that the proposed end-to-end HIL design framework provides steps forward to mitigate effect of uncontrolled artifacts and mismatch between mathematical modeling and real-life experiments. It is worth mentioning that programmable phase-only SLMs are not compact, light and cheap, contrary to it, they are bulky and very expensive. However, the designed optimal SLM phase-pattern can be used as desirable phase profile for manufacturing of hardware DOEs such as diffractive and meta optical elements.

2 RELATED WORK

As relevant to this work, we provide references to achromatic EDoF imaging with DOEs and the hardware-in-the-loop technique as applied to imaging problems.

Achromatic EDoF imaging. EDoF with minimal chromatic aberrations (see Figure 3) is a classic problem of computational imaging that has been intensively studied in recent works motivated by development of new lighter, and compact optical systems for emerging imaging applications such as computational photography. The reason to study this problem in modern times is because classical compound multi-array of lenses solutions do not either apply [Tseng et al. 2021] or extend the DoF for the new compact optical systems.

One of the basic tools of all-focus imaging is an optics with depth invariant Point Spread Function (PSF) referring to the prominent wavefront coding (WFC) proposed in [Dowek and Cathey 1995]. A flow of publications exploiting this idea in different optical setups and for various goals. The recent works develop this idea for imaging with diffractive [Sitzmann et al. 2018] and meta-plates [Bayati et al. 2021; Colburn et al. 2018; Tseng et al. 2021]. The designed depth invariant PSFs serve as the main components of inverse imaging. For inverse imaging in physical experiments, the calculated PSFs are calibrated and corrected according to the corresponding test for manufactured DOEs. In this way, a gap between theoretical and physical image formation can be diminished following the logic: first, design of DOE and, after, its tuning according to the experimental study of the manufactured DOEs.

The sensitivity of inverse imaging with respect to errors in PSF design and calculation can be compensated to some extent by using convolutional neural network (CNN) technologies [Zeng and Lam 2021] dealing directly with the blurred images and designing the inverse imaging algorithms based on test-images as it is particularly done in the works [Alghamdi et al. 2021; Chen et al. 2020a; Liu et al. 2021; Sun et al. 2021; Tseng et al. 2021]. Note that in our approach based on HIL-SLM setup this problem is eliminated completely, since we do not deal with PSFs at all as we do not need them for SLM phase-pattern design and do not use them for inverse imaging.

In this work the elements of interest to be jointly designed are refractive lens and DOE (hybrid optics) as it is shown in Figure 3 to improve DoF and reduce chromatic aberrations of a system. Hybrid optics for achromatic EDoF appeared in the works [Flores et al. 2004; Liu 2007; Liu et al. 2007] where it was designed as a focusing lens with direct evaluation of optical performance. In [Rostami et al. 2021], design in terms of end-to-end optimization was developed targeting phase-encoded inverse imaging with modification of the hybrid to optimal sharing of optical power between DOE and lens.

Hardware-in-the-loop. The HIL approach to optimization is known well and for a long time in engineering and science. Actually, any adjustment of hardware when we change something and then look at the result can be treated as the HIL procedure. The main problem is that parameters to be optimized should be easy for variations.

We know only a few works concerning application of HIL for imaging problems close to being considered in this paper. In the work [Peng et al. 2020], HIP approach is used to fit parameterized wave propagation model using the phase-only SLM, which results in a dramatically improved performance of the holographic display. In [Tseng et al. 2019], HIP setup is used for tuning of ISP parameter provided fixed optics. To the best of our knowledge, what is done in this paper is a first attempt to use HIL in design of DOEs for intensity imaging.

3 OPTIMIZATION-IN-THE-LOOP OF HYBRID OPTICS

The proposed design framework of hybrid optics for achromatic EDoF is summarized in Algorithm 1 which follows an alternating methodology: fixing hyperparameter $\Theta_{SLM}$, solving $\Theta_{opt}$ for inverse imaging, updating $\Theta_{SLM}$ and so forth. We implement this...
which is then passed to the downstream reconstruction (deblurring) module. The output of the task module is evaluated using trained CNNs and a weighted combination of PSNR between estimated and ground truth images, \( \mathcal{L}_{PSNR} \), and perceptual losses \( \mathcal{L}_{Adv} \) and \( \mathcal{L}_{Percept} \), with weights \( \sigma_1, \sigma_2, \text{ and } \sigma_3 \).

**Algorithm 1: HIL Design of SLM Phase-pattern and Inverse Imaging**

**Input (\( A, N_{iter} \))**

- Initialize CMA-ES, \( \Theta_{SLM} \leftarrow \Theta_{SLM}^{(0)}, t \leftarrow 1; \)
- Train initial inverse imaging CNN for \( \Theta_{SLM} \) for \( 0.5, 1.0, 1.8m \).
- \( \Theta_{soft} \leftarrow \Theta_{soft}^{(0)} \) for a wide range of hyperparameters \( \Theta_{SLM} \);
- while \( t \leq N_{iter} \) do
  - for \( r = 1 \) to \( R \) do
    - \( \Theta_{SLM}^{(r)} \leftarrow \) randomly draw from Gaussian at \( \Theta_{SLM} \);
    - Add random noise to \( \Theta_{SLM}^{(r)} \);
    - \( s\left( \Theta_{SLM}^{(r)} \right) \) get blurred data at \( 0.5, 1.0, 1.8m \) of images \( I_1, \ldots, I_N \) and use \( \Theta_{soft}^{(r)} \) CNN to estimate them.
    - \( \mathcal{L}_{HIL}\left( s\left( \Theta_{SLM}^{(r)} \right) \right) \) compute average PSNR among the estimated \( f\)-images at each distance 0.5, 1.0, 1.8m.
    - \( \Theta_{SLM} \leftarrow \) update CMA-ES.
  - \( t \leftarrow t + 1; \)
  - \( \Theta_{soft}^{(r)} \) train inverse imaging CNN for best SLM-pattern among \( \Theta_{SLM}^{(1)}, \ldots, \Theta_{SLM}^{(R)} \);
- return \( \Theta_{SLM}, \Theta_{soft}; \)

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 \( \Theta_{SLM}^{(0)} \) and a pre-trained inverse imaging CNN for a wide range of hyperparameter \( \Theta_{SLM} \). After this, the 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, \( s\left( \Theta_{SLM}^{(r)} \right) \), 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), \( \mathcal{L}_{HIL}\left( s\left( \Theta_{SLM}^{(r)} \right) \right) \). Then, using the \( 0th\)-order stochastic evolutionary search method CMA-ES\(^1\) [Hansen and Ostermeier 1996], Algorithm 1 updates \( \Theta_{SLM} \) by 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 \( \Theta_{SLM}^{(r)} \). Performing the previous alternating process \( N_{iter} \) times algorithm returns the updated \( \Theta_{SLM} \) and \( \Theta_{soft} \). The structure of CNN developed for inverse imaging (optimization on \( \Theta_{soft} \)) is shown in Figure 4. Algorithm 1 can be also initiated by solutions obtained according to the model-based approach from [Rostami et al. 2021]. The number of ‘global’ iterations of this algorithm for \( \Theta_{soft} \) is \( N_{iter} \) and \( R \) is the number of ‘local’ iterations for \( \Theta_{SLM} \). In this work we fixed \( N_{iter} = 3 \) and \( R = 500 \). In the following sections, more details per each stage are described.

3.1 Parameterized Optics Model

The phase-profile of SLM is designed as a piece-wise spatially invariant function defined for the design wavelength \( \lambda_0 \). Following to [MiriRostami et al. 2021], we start from the absolute phase model which further is wrapped to the interval defined by the modulation range of SLM and discretized with a number of the parameters to be optimized equal to fourteen which we introduce in brief. The proposed absolute phase \( \varphi_{\lambda_0} \) takes the form:

\[
\varphi_{\lambda_0}(x, y) = \frac{2\pi}{\lambda_0 f_{b_0}} \left( x^2 + y^2 \right) + \beta(x^3 + y^3) + \sum_{p=1, p \neq 4}^{P} \rho_p Z_p(x, y),
\]

(1)

where the first term models the squared phase of the lens which fits the form of the fourth Zernike polynomial omitted in the sum of the Zernike components, \( c \) encapsulates the focusing contribution of the SLM for wavelength \( \lambda_0 \), and the lens focal length \( f_{b_0} \) fixed as \( f_{b_0} = 0.01m \) in this work. The cubic phase of magnitude \( \beta \) is a typical component for EDoF as introduced in [Dowski and Cathey 1995], the third group of the items is for parametric approximation of the free-shape DOE using the Zernike polynomials \( Z_p(x, y) \) with coefficients \( \rho_p \) to be estimated. Thus, the full set of parameters to be optimized for SLM is defined as \( \Theta_{SLM} = (\alpha, \beta, \rho_1, \ldots, \rho_P) \).

Fresnel Order (Thickness). The SLM phase profile in radians is defined as \( \Theta_{SLM} = 2\pi m_Q \), where \( m_Q \) is called ‘Fresnel order’ of the

\(^1\)documentation in python of the CMA-ES optimizer in https://pypi.org/project/cma/.
where ‘Scene’ denotes objects under investigation; the polarizer, ‘P’, keeps the light polarization needed for a proper wavefront modulation by SLM; the beamsplitter, ‘BS’, governs SLM illumination and further light passing; the lenses ‘L₁’ and ‘L₂’ form a 4f-telescopic system transferring the light wavefront modified by SLM to the lens ‘L₃’ plane; the lens ‘L₃’ forms an image of the ‘scene’ on the imaging detector, ‘CMOS’. Since the main goal of this paper is to extend the DoF of a photography camera we have three scenes located at different depths which are physically implemented by three identical polarized 15.6° monitors².

Test images in ‘scene’ plane are displayed on these monitors with 1920 × 1080 pixels and 570 ppi. The distances d₁ for the monitors are 0.5, 1.0, 1.8 m. The image dataset we use as scenes consists of 1000 high-resolution images³. We point out that the image sizes at each monitor (depths 0.5, 1.0, 1.8 m) are scaled in such a way that the sizes of the corresponding registered images at the sensor are the same, in sensor pixel 512 × 512.

For DOE implementation, we use the Holocue phase-only GAEA-2-vis SLM panel, resolution 4160 × 2464, pixel size 3.74 μm; ‘L₁’ and ‘L₂’ achromatic doublet lenses with diameter 12.7 mm and focal distance of 50 mm; BK7 glass lens ‘L₃’ with diameter 9.2 mm and focal distance 10.0 mm (approximately d₂); ‘CMOS’ Blackfly S board Level camera with the color pixel matrix Sony IMX264, 3.45 μm pixels and 2448 × 2048 pixels. SLM allows us to study the hybrid optics with the phase distribution of the designed DOE (implemented on SLM) additive to the imaging lens ‘L₃’. The DOE phase was created as an 8-bit .bmp file and imaged on SLM. We calibrated the SLM phase delay response to the maximum value of 3.6π (fresnel order equal to 1.8) for a wavelength of 510 nm. This 3.6π corresponds to the value 255 of .bmp file for the phase image of DOE.

3.3 Inverse Imaging: Update for Θ_{soft}
Data processing here solves the following problems: 1) inverse imaging (reconstruction of sharp images from the registered blurred ones); 2) compensation of various errors such as phase modulation errors in SLM, errors in demosaicing software, noise in CMOS, etc. The learning NN approach within an end-to-end design is used to resolve these problems. In this work we implement a DRUNet CNN architecture [Zhang et al. 2021] illustrated in Figure 4. We remark that this network has the ability to handle various noise levels for an RGB image, per channel, via a single model. The backbone of DRUNet is U-Net which consists of four scales. Each scale has an identity skip connection between 2 × 2 strided convolution (SConv) downsampling and 2 × 2 transposed convolution (TConv) upsampling operations. The number of channels in each layer from the first scale to the fourth scale are 64, 128, 256 and 512, respectively. Four successive residual blocks are adopted in the downsampling and upsampling of each scale. Each residual block only contains one ReLU activation function. It is worth noting that the proposed DRUNet is biasfree, which means no bias is used in all the Conv, SConv and TConv layers [Zhang et al. 2021].

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, L_{PSNR}, and perceptual losses given below by:

$$L_{PSNR} = -10 \log_{10} \frac{\sum_{i,j} (I_{ref}(i,j) - I_{est}(i,j))^2}{\sum_{i,j} I_{ref}(i,j)^2}$$

$$L_{perceptual} = \sum_{i,j} \| \nabla I_{ref}(i,j) - \nabla I_{est}(i,j) \|^2$$

³Blackstorm MobileMonitor Their Description can be found in https://www.verkkokauppa.com/fi/product/14167/mxvvq/Blackstorm-MobileMonitor-15-6-Full-HD-kanettava-naytto.

²Image databases employed in the training stage can be found in https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/EECTVN.
We propose a nonlinear black-box optimizer that uses CMA-ES strat-
VGG-16 [Simonyan and Zisserman 2014] model for our perceptual
values of about 25 dB for different imaging depths and colors is achieved by the designed hybrid. In contrast, while the lens-only system performs well for the
infocus distance $d_1 = 1.0$ m and the imaging quality degrade essentially for the off-focus distances.

**Perceptual loss:** To measure the semantic difference between
the estimated output and the ground truth, we use a pretrained
VGG-16 [Simonyan and Zisserman 2014] model for our perceptual
loss [Khan et al. 2020]. We 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 $L_{\text{Percep}}$ is the averaged PSNR
between the outputs of these two activation functions for both
estimated and ground truth images.

**Adversarial loss:** Adversarial loss [Goodfellow et al. 2014] was
added to further bring the distribution of the reconstructed output
close to those of the real images. Given the swish activation function
[Ramachandran et al. 2017] as our discriminator $D$, this loss is given as $L_{\text{Adv}} = -\log(D(I_{\text{test}}))$ where $I_{\text{test}}$ 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,$L_{\text{CNN}} = \sigma_1 L_{\text{PSNR}} + \sigma_2 L_{\text{Percep}} + \sigma_3 L_{\text{Adv}}$, 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 $\Theta_{\text{soft}}$.

3.4 HIL Optimizer: Update for $\Theta_{\text{SLM}}$

We propose a nonlinear black-box optimizer that uses CMA-ES stra-

**HIL Optimizer:** Update for $\Theta_{\text{SLM}}$.

We propose a nonlinear black-box optimizer that uses CMA-ES stra-

**HIL Optimizer:** Update for $\Theta_{\text{SLM}}$.

4 EXPERIMENTAL RESULTS

In this section we evaluate and discuss the imaging results of the
developed hybrid system over the testing dataset. These results are shown in Figures 6 and 7, respectively, for Setups 1 and 2 as
noted in Figure 1. Remind that in Setup 1 the three monitors display
images of the three distances from the sensor. In this scenario, we
are able to evaluate the quality of reconstructions visually as well
as numerically by PSNR values for each of the RGB color channels.

Setup 2 is developed to evaluate EDoF where the scene is composed
by flowers located at different distances from sensor to validate the
performance of our design in real-world scenario. In this case, the
quality of imaging can be evaluated only visually (qualitatively).

**Setup 1** The results shown in Figure 6 are presented in three rows,
respectively, for three depth distances (object-SLM): $d_1 = 0.5$ m
(row 1), $d_1 = 1.0$ m (row 2), and $d_1 = 1.8$ m (row 3). The groups
of columns Groundtruth, Designed hybrid and Lens-only show the
true images as displaced by the three monitors for different depths,
reconstruction by the developed algorithm, and reconstruction by
the lens-only system. In the last case, the SLM is switched off (no
phase modulation) and imaging is produced by the lens only. For
a fair comparison, the optimization of inverse imaging, in this case, is
produced identically to those for the hybrid system. The experiments
with the lens-only are produced in order to evaluate the effects of
the phase-modulation by SLM. In the columns Designed hybrid and
Hybrid Diffractive Optics Design via Hardware-in-the-Loop Methodology for Achromatic Extended-Depth-of-Field Imaging

Fig. 7. Setup 2. Comparison of the diffractive optics versus the compound lens Sony A7 III (85mm focal length with F22) and iPhone Xs Max cameras. The obtained images are presented in row 3 with their enlarged fragments in rows 4, 5, 6 corresponding to three off-focus distances $d_1 = 0.4, 0.65, 1.9$ m, respectively. For comparison also included results obtained by lens-only and lens + SLM of cubic phase-pattern. The visual advantage in sharpness and color preservation is clearly in favor of the designed hybrid. The images of the sunflower moving in the $d_1$ depth range 1.9-0.3 m are shown in column 6. The sharpness and color presentation are very good despite variations of distance.

**Lens-only**, we can see blurred images used for reconstruction as well as the zoomed fragments of the reconstructions shown also for the true images. The zoomed sections for blurry and reconstructed images visually reveal clearly that the lens system suffers from strong chromatic aberrations and the quality of imaging is lost especially for defocus distances $d_1 = 0.5$ m and $d_1 = 1.8$ m. The best performance for the both compared system is demonstrated for the focal distance $d_1 = 1.0$ m. However, the advantage of the hybrid is well seen in qualitative and numerical comparison. For the defocus distances the advantage of the designed hybrid is much more valuable. For instance, for $d_1 = 1.8$ m, the improvement of about 2dB is in favor of the hybrid. Moreover, for the designed hybrid, the PSNR values for different colors and depths are more or less the same at about 25 dB. It confirms that the designed hybrid imaging indeed demonstrates achromatic EDoF imaging.

**Setup 2.** In this setup, the whole 3D scene is observed. It is the all-in-focus scenario when the goal is to get in-focus images for all object of different distances from SLM. Figure 7 shows the performance for: the designed hybrid (column 5), lens + cubic absolute...
phase (column 4), lens-only (column 3), and two compound commercial cameras, Sony A7 III (column 1) and iPhone Xs Max (column 2) cameras. In this experiment, 6 flowers are located at the different distances $d_1 = 0.4$ m (cherry blossom), 0.65 m (first red rose), 1.0 m (tulip flower), 1.6 m (second red rose), 1.8 m (sunflower), and 1.9 m (magnolia flower) as it is shown in Figure 1. It is worth mentioning that for Sony and iPhone (compound optics), we adjusted the focusing point to $d_1 = 1.0$ m (tulip flower) for a fair comparison with other imaging systems. For infocus point ($d_1 = 1$ m), Sony camera produces the best result but with a narrowed DoF. On the other hand, the imaging quality for iPhone camera is not good for distances less than 1.0 m.

The imaging results are presented in row 3. For more detailed comparison, the three zoom fragments of these images corresponding to different distances $d_1$ are shown in rows 4, 5, and 6. The visual advantage of the hybrid in both the sharpness of imaging and the proper color preservation is quite obvious. The lens-only and cubic hybrid setups exhibit strong chromatic aberrations with less imaging accuracy for both focus and out-of-focus situations similar as it was seen in Figure 6. The most important in these experiments is a comparison of our designed hybrid versus the compound lens Sony and iPhone cameras, where the designed hybrid clearly shows better all-in-focus imaging. In addition to these experiments, we acquired a video of the sunflower moving in the depth range (1.9–0.3 m). A few frames of this video are presented in column 6 of Figure 7. One may note that despite of the distance variations the sunflower imaging is sharp with good color preservation for all distances. Note, that the shown frames are given for distances different from those used in optimization. In summary, these experiments suggest the effectiveness of the HIL-SLM setup as a base methodology for designing diffractive imaging systems.

5 CONCLUSION
It is shown in this paper that the optimized hybrid optical system composed from refractive lens and DOE in the scenario of achromatic EDoF imaging demonstrates advanced performance as compared with the single refractive lens, lens + cubic absolute phase component, and two compound commercial cameras, iPhone Xs Max and Sony A7 III cameras. In experiments, the hybrid optics is implemented by optical projection of the SLM phase pattern on a lens plane for a wide wavelength interval (400–700 nm) and the depth-of-field range (0.4–1.9 m). Multiple results comparing the developed DOE imaging system with advanced conventional compound multi-lens cameras such as Sony A7 III and iPhone Xs Max cameras show its competitive imaging quality and advanced in all-focus sharp imaging.

REFERENCES
Mashael Alghamdi, Qiang Fu, Ali Thabet, and Wolfgang Heidrich. 2021. Transfer Deep Learning for Reconfigurable Snapshot HDR Imaging Using Coded Masks. In Computer Graphics Forum. Wiley Online Library.
Nick Antipa, Grace Kuo, Reinhard Heckel, Ben Mildenhall, Emrah Rostam, Ren Ng, and Laura Waller. 2018. DiffuserCam: lensless single-exposure 3D imaging. Optica 5, 1 (2018), 1–9.
Henry Arguello, Samuel Pinilla, Yifan Peng, Hayato Ikoma, Jorge Bacca, and Gordon Wetzstein. 2017. Shift-variant color-coded diffractive spectral imaging system. Optica 8, 11 (2021), 1424–1434.
M. Salman Asif, Ali Ayremlou, Aswin Sankaranarayanan, Ashok Veeraraghavan, and Richard G. Baraniuk. 2017. FlatCam: Thin, Lensless Cameras Using Coded Aperture and Computation. IEEE Transactions on Computational Imaging 3, 3 (2017), 384–397. https://doi.org/10.1109/TCI.2016.2593662
Anne Auger and Nikolaus Hansen. 2012. Tutorial CMA-ES: evolution strategies and covariance matrix adaptation. In Proceedings of the 14th annual conference companion on Genetic and evolutionary computation. 827–848.
Seung-Eun Baek, Hayato Ikoma, Daniel S Jemyn, Yuqi Li, Wolfgang Heidrich, Gordon Wetzstein, and Min H Kim. 2020. End-to-End Hyperspectral-Depth Imaging with Learned Diffractive Optics. arXiv preprint arXiv:2009.04630 (2020).
Elyas Bayati, Raphael Pestourie, Shane Colburn, Zin Lin, Steven G. Johnson, and Arka Majumdar. 2021. Inverse designed extended depth of focus meta-optics for broadband imaging in the visible. Nanophotonics (2021), 0000101515202104531. https://doi.org/10.1515/nanoph-2021-0431
Vivek Boominathan, Jesse K. Adams, M. Salman Asif, Benjamin W. Avants, Jacob T. Robinson, Richard G. Baraniuk, Aswin C. Sankaranarayanan, and Ashok Veeraraghavan. 2016. Lensless Imaging: A computational renaissance. IEEE Signal Processing Magazine 33, 5 (2016), 23–35. https://doi.org/10.1109/MSP.2016.2581921
Julie Chang, Vincent Sitzmann, Xiong Dun, Wolfgang Heidrich, and Gordon Wetzstein. 2018. Hybrid optical-electronic-convolutional neural networks with optimized diffractive optics for image classification. Scientific reports 8, 1 (2018), 1–10.
Peidong Chen, Xuqin Su, Muyuan Liu, and Wenhui Zhu. 2020a. Lensless computational imaging technology using deep convolutional network. Sensors 20, 9 (2020), 2661.
Wei Ting Chen, Alexander Y Zhu, and Federico Capasso. 2020b. Flat optics with dispersion-engineered metasurfaces. Nature Reviews Materials 5, 8 (2020), 604–620.
Shane Colburn and Arka Majumdar. 2021. Inverse design and flexible parameterization of meta-optics using algorithmic differentiation. Communications Physics 4, 1 (2021), 1–11.
Shane Colburn, Alan Zhan, and Arka Majumdar. 2018. Meta Surfaces for full-color computational imaging. Science Advances 4, 2 (2018), eaar2114. https://doi.org/10.1126/sciadv.aar2114
Edward R. Dowek and W. Thomas Cathey. 1995. Extended depth of field through wave-front coding. Appl. Opt. 34, 11 (Apr 1995), 1859–1866. https://doi.org/10.1364/AO.34.018559
Paul Dudoshé and Nikolaus Hansen. 2021. Augmented Lagrangian, penalty techniques and surrogate modeling for constrained optimization with CMA-ES. In Proceedings of the Genetic and Evolutionary Computation Conference (GECCO).
Xiong Dun, Hayato Ikoma, Gordon Wetzstein, Zhanhuang Wang, Xinhin Cheng, and Yifan Peng. 2020. Learned rotationally symmetric diffractive achromat for full-spectrum computational imaging. Optica 7, 8 (Aug 2020), 913–922. https://doi.org/10.1364/OPTICA.394413
Jacob Engelberg and Uriel Levy. 2020. The advantages of meta-lenses over diffractive lenses. Nature Communications 11, 1 (2020), 1–4.
Angel Flores, Michael R. Wang, and Jane Y. Wang. 2004. Achromatic hybrid refractive-diffractive lens with extended depth of focus. Applied optics 43, 30 (2004), 5516–5530.
Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. Advances in neural information processing systems 27 (2014).
Nikolaus Hansen and Andreas Ostermeier. 1996. Adapting arbitrary normal mutation distributions in evolution strategies: The covariance matrix adaptation. In Proceedings of IEEE international conference on evolutionary computation. IEEE, 312–317.
Salim Siddique Khan, Varun Sundar, Vivek Boominathan, Ashok Veeraraghavan, and Kaushik Mitra. 2020. Flatnet: Towards photorealistic scene reconstruction from lensless measurements. IEEE Transactions on Pattern Analysis and Machine Intelligence (2020).
Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014).
Yaukun Liu, Chongyang Zhang, Tingdong Kou, Yueyi Li, and Junfei Shen. 2021. End-to-end computational optics with a single lens for large depth-of-field imaging. Optics Express 29, 18 (2021), 28530–28548.
Zhiqiang Liu. 2007. Diffractive lens with extended depth of focus and its applications. Ph.D. Dissertation. University of Miami.
Zhiqiang Liu, Angel Flores, Michael R. Wang, and Jianwen J. Yang. 2007. Diffractive infrared lens with extended depth of focus. Optical Engineering 46, 1 (2007), 1–9. https://doi.org/10.1117/1.2430506
SeyyedReza MiriRostami, Vladimir Y. Katkovnik, and Karen O. Eguiazarri. 2021. Extended DoF and achromatic inverse imaging for lens and lensless MPM camera based on Wiener filtering of defocused OTFs. Optical Engineering 60, 5 (2021), 1–14. https://doi.org/10.1117/1.OE.60.5.051204
Krisztina Monakhova, Kyrolos Yanny, Neerja Aggarwal, and Laura Waller. 2020. Spectral DiffuserCam: lensless snapshot hyperspectral imaging with a spectral filter array. Optica 7, 10 (Oct 2020), 1298–1307. https://doi.org/10.1364/OPTICA.397214
Ali Mesleh, Avinash Sharma, Emmanuel Onzon, Fahim Mannan, Nicolas Robolbux, and Tolkf Heide. 2020. 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).
Yifan Peng, Suyeon Choi, Nitsul Padmanaban, and Gordon Wetzstein. 2020. Neural holography with camera-in-the-loop training. ACM Transactions on Graphics (TOG)
Hybrid Diffractive Optics Design via Hardware-in-the-Loop Methodology for Achromatic Extended-Depth-of-Field Imaging

Prajit Ramachandran, Barret Zoph, and Quoc V Le. 2017. Searching for activation functions. arXiv preprint arXiv:1710.05941 (2017).

Seyyed Reza Miri Rostami, Samuel Pimilla, Igor Shevkunov, Vladimir Katkovnik, and Karen Egiazarian. 2021. Power-balanced hybrid optics boosted design for achromatic extended depth-of-field imaging via optimized mixed OTF. Appl. Opt. 60, 30 (Oct 2021), 9365–9378. https://doi.org/10.1364/AO.434852

Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014).

Vincent Sitzmann, Steven Diamond, Yifan Peng, Xiong Dun, Stephen Boyd, Wolfgang Heidrich, Felix Heide, and Gordon Wetzstein. 2018. End-to-End Optimization of Optics and Image Processing for Achromatic Extended Depth of Field and Super-Resolution Imaging. 37, 4, Article 114 (Jul 2018), 13 pages. https://doi.org/10.1145/3197517.3201333

Qilin Sun, Congli Wang, Fu Qiang, Dun Xiong, and Heidrich Wolfgang. 2021. End-to-end complex lens design with differentiable ray tracing. ACM Trans. Graph. 40, 4 (2021).

Jasper Tan, Li Niu, Jesse K. Adams, Vivek Boominathan, Jacob T. Robinson, Richard G. Baraniuk, and Ashok Veeraraghavan. 2019. Face Detection and Verification Using Lensless Cameras. IEEE Transactions on Computational Imaging 5, 2 (2019), 180–194. https://doi.org/10.1109/TCI.2018.2889933

Ethan Tseng, Shane Colburn, James Whitehead, Luocheng Huang, Seung-Hwan Baek, Arka Majumdar, and Felix Heide. 2021. Neural nano-optics for high-quality thin lens imaging. Nature Communications 12, 1 (29 Nov 2021), 6493.

Ethan Tseng, Felix Yu, Yuting Yang, Fahim Mannan, Karl St. Arnaud, Derek Nowrouzezahrai, Jean-Francois Lalonde, and Felix Heide. 2018. Hyperparameter Optimization in Black-box Image Processing using Differentiable Proxies. ACM Transactions on Graphics (TOG) 38, 4 (Jul 2019). https://doi.org/10.1145/3306346.3322996

Tianjiao Zeng and Edmund Y. Lam. 2021. Robust Reconstruction With Deep Learning to Handle Model Mismatch in Lensless Imaging. IEEE Transactions on Computational Imaging 7 (2021), 1080–1092. https://doi.org/10.1109/TCI.2021.3114542

Kai Zhang, Yawei Li, Wangmeng Zuo, Lei Zhang, Luc Van Gool, and Radu Timofte. 2021. Plug-and-play image restoration with deep denoiser prior. IEEE Transactions on Pattern Analysis and Machine Intelligence (2021).