https://doi.org/10.2352/EI.2023.35.9.IPAS-287

Paper: Hybrid diffractive optics (DOE & refractive lens) for broadband EDoF imaging

Seyyed Reza Miri Rostami, Samuel Pinilla, Igor Shevkunov, Vladimir Katkovnik, and Karen Egiazarian; Computing Sciences Unit, Faculty of Information Technology and Communication Sciences, Tampere University, FI-33720 Tampere, Finland

Abstract

In the considered hybrid diffractive imaging system, a refractive lens is arranged simultaneously with a multilevel phase mask (MPM) as a diffractive optical element (DOE). Extended depth-of-field (EDoF) imaging and low chromatic aberrations are the two potential advantages of the proposed hybrid setup. To make use of these advantages, this paper proposes a fully differentiable image formation model that uses neural network techniques to maximize the imaging quality by optimizing MPM, digital image reconstruction algorithm, refractive lens parameters (aperture size, focal length) and distance between the MPM and sensor. In the first stage of the design framework, model-based numerical simulations and end-to-end joint optimization of imaging are used. A spatial light modulator (SLM) is employed at the second stage of the design to implement MPM optimized at the first stage, and the image processing is optimized experimentally using a learning-based approach. The third stage of optimization is targeted at joint optimization of the SLM phase pattern and image reconstruction algorithm in the hardware-in-the-loop (HIL) setup, which allows compensating a mismatch between numerical modeling and physical reality of optics and sensor. A comparative analysis of the imaging accuracy and quality using the aforementioned optical parameters is presented. For the first time, varying aperture sizes, lens focal lengths, and distances between MPMs and sensors for end-to-end optimization of EDoF is considered. The numerical and experimental comparisons are performed between the designs for the visible wavelength interval [400-700] nm and the following EDoF ranges for simulations and experiments [0.5-100] m and [0.5-2.0] m, respectively. Using SLM as a programmable DOE allows to study the potential of imaging with wavefront phase modulation. It is proved experimentally, first time to the best of our knowledge, that wavefront phase modulation is able to provide imaging of advanced quality as compared with some commercial multi-lens cameras.

Introduction

End-to-end optimization of diffractive optical element (DOE) profile (e.g., binary/multi-level phase elements [1–3]; meta-optical elements included [4–8]) has gained increasing attention in emerging applications such as photography [9, 10], augmented reality [11], spectral imaging [12], microscopy [13], among others that are leading the need for highly miniaturized optical systems [14, 15], etc. As part of the design methodology, numerical differentiable model is built for propagation of light fields through the physical setup in order to be used for modeling and optimization methods employing neural networks. Particularly, [16] proposes and studies the power-balanced diffractive hybrid optics (lens and MPM), which is the methodology that will be used in this study, as it entails the use of spatial light modulators to encode light fields using MPM.

In this work, the elements of interest to be jointly designed are MPM and image processing algorithms. The techniques and algorithms used for this design take advantage of those developed in [16, 17]. As in [16], the targeted imaging problem is Extended Depth-of-Field (EDoF) with reduced chromatic aberrations. We exploit a fully differentiable image formation model for joint optimization of optical and imaging parameters for the designed computational camera using neural networks. In particular, for the number of levels and Fresnel order features, we introduce a smoothing function because both parameters are modeled as piecewise continuous operations. As an alternative approach, to bridging the gap between the numerical solution and real-world physical implementation, we implemented end-to-end design of the SLM pattern as DOE in optical setup through a ”hardware-in-the-loop (HIL)” imaging setup, for achromatic EDoF RGB imaging. In this case, we followed the mainstream of the design proposed in [17]. Improvements in imaging is demonstrated as in iteration of optimization as well as in comparison of different DOE designs: model-based and HIL-based.

This paper is an extension and further development of our conference paper [18]. It proves experimentally (section), first time to the best of our knowledge, that wavefront phase modulation is able to provide imaging of competitive quality as compared with some commercial compound multi-lens cameras.

In this paper, we focus on practical aspects of design, especially on the imaging quality and accuracy as a function of basic optical parameters: aperture size, lens focal length, MPM thickness, the distance between MPM and sensor, and the F-number. The designed systems are compared numerically and experimentally for the wavelength range of $(400-700)$ nm and depth-of-field range of $(0.5-100)$ m and (0.5-2) m, respectively. The study concerns the application of hybrid optics for compact cameras with aperture $(5-9)$ mm and lens focal length $(3-10)$ mm. We point out that the variables aperture size, lens focal length, and distance between MPM and sensor are for the first time considered for end-to-end optimization of EDoF.

The contribution of this work can be summarized as follows.

- End-to-end optimization methodology for the joint design of DOE, parameters of refractive lens and imaging algorithms, showing high efficiency in terms of image accuracy and visual quality;
- Optimal hybrid setup in terms of the optimal balance between aperture size and lens focal length concluded from multiple simulated experiments;
- Algorithms for using SLM as MPM in the hybrid optics with

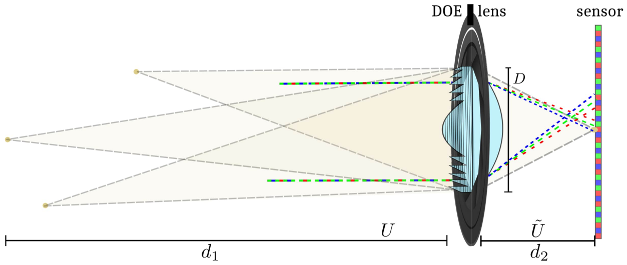

Figure 1. *A light wave with a given wavelength and a curvature for a point source at a distance d_1 propagates to the aperture plane containing MPM (refractive index n) to be designed. The MPM modulates the phase of the incident wavefront. Using Fresnel propagation, the resulting wavefront propagates through the lens to the aperture sensor, distance d_2. The intensities of the sensor-incident wavefront define PSFs of the diffractive hybrid optical system.*

learning-based CNN optimization of inverse imaging ;

- Showing the advance performance of HIL methodology for co-design of SLM phase-pattern and inverse imaging for achromatic EDoF compared with the model-based approach;
- The advanced achromatic EDoF imaging of the designed system as compared with conventional compound multi-lens cameras such as in smartphones.

Model-based End-to-End Optimization of Imaging with Hybrid optics

The optical setup is depicted in Figure 1, where object, aperture, and sensor are $2D$ flat, d_1 is a distance between the object and the aperture, d_2 is a distance from the aperture to the sensor ($d_2 \ll d_1$), f_{λ_0} is a lens focal length. In what follows, we use coordinates (x,y), and (u,v) for aperture, and sensor planes, respectively. In this section, we mainly follow the image formation modeling and design optimization presented in [16]. These results are included for the completeness of the presentation and in order to give a clear picture of our approach, methodology, and algorithms.

Image Formation Model

PSF-based RGB imaging

Based on the Fresnel diffraction wavefront propagation, the response of an optical system to an input wavefront is modeled as a convolution of the system's PSF and a true object-image. Let us assume that there are both a lens and MPM in the aperture, then a generalized pupil function of the system for intensity imaging shown in Figure 1 is of the form (see Eqs. (5-23)-(5-28) in [19])

$$\mathscr{P}_\lambda(x,y) = \mathscr{P}_A(x,y)e^{\frac{j\pi}{\lambda}\left(\frac{1}{d_1}+\frac{1}{d_2}-\frac{1}{f_\lambda}\right)(x^2+y^2)+j\varphi_{\lambda_0,\lambda}(x,y)}. \quad (1)$$

In (1), f_λ is a lens focal length for the wavelength λ, $P_A(x,y)$ represents the aperture of the optics and $\varphi_{\lambda_0,\lambda}(x,y)$ models the phase delay enabled by MPM for the wavelength λ provided that λ_0 is the wavelength design-parameter for MPM. In this formula, the phase $\frac{j\pi}{\lambda}\left(\frac{1}{d_1}+\frac{1}{d_2}\right)(x^2+y^2)$ appears due to propagation of the coherent wavefront from the object to the aperture (distance d_1) and from the aperture to the sensor plane (distance d_2), and $\frac{-j\pi}{\lambda f_\lambda}(x^2+y^2)$ is a quadratic phase delay due to the lens. For the lensless system

$$\mathscr{P}_\lambda(x,y) = \mathscr{P}_A(x,y)e^{\frac{j\pi}{\lambda}\left(\frac{1}{d_1}+\frac{1}{d_2}\right)(x^2+y^2)+j\varphi_{\lambda_0,\lambda}(x,y)}, \quad (2)$$

and for the lens system without MPM, $\varphi_{\lambda_0,\lambda}(x,y) \equiv 0$ in (1).

In the hybrid system, which is the topic of this paper, the generalized aperture takes the form

$$\mathscr{P}_\lambda(x,y) = \mathscr{P}_A(x,y)e^{\frac{j\pi}{\lambda}\left(\frac{1}{d_1}+\frac{1}{d_2}-\frac{1}{f_\lambda}\right)(x^2+y^2)+j\varphi_{\lambda_0,\lambda,\alpha}(x,y)}, \quad (3)$$

where the optical power of the hybrid is shared between the lens with the optical power $1/f_\lambda$ and the MPM due to the quadratic phase component included in the phase delay of MPM. The magnitude of the latter phase is controlled by a real-valued parameter α.

The PSF of the coherent monochromatic optical system for the wavelength λ is calculated by the formula [19]

$$PSF_\lambda^{coh}(u,v) = \mathscr{F}_{\mathscr{P}_\lambda}\left(\frac{u}{d_2\lambda}, \frac{v}{d_2\lambda}\right), \quad (4)$$

where $\mathscr{F}_{\mathscr{P}_\lambda}$ is the Fourier transform of $\mathscr{P}_\lambda(x,y)$. Then, PSF for the corresponding incoherent imaging, which is a topic of this paper, is a squared absolute value of $PSF_\lambda^{coh}(u,v)$. After normalization, this PSF function takes the form

$$PSF_\lambda(u,v) = \frac{\left|PSF_\lambda^{coh}(u,v)\right|^2}{\iint_{-\infty}^{\infty}\left|PSF_\lambda^{coh}(u,v)\right|^2 dudv}. \quad (5)$$

We calculate PSF for RGB color imaging assuming that the incoherent radiation is broadband and the intensity registered by an RGB sensor per c-band channel is an integration of the monochromatic intensity over the wavelength range Λ with the weights $T_c(\lambda)$ defined by the sensor color filter array (CFA) and spectral response of the sensor. Normalizing these sensitivities on λ, i.e. $\int_\Lambda T_c(\lambda)d\lambda = 1$, we obtain RGB channels PSFs

$$PSF_c(u,v) = \frac{\int_\Lambda PSF_\lambda(u,v)T_c(\lambda)d\lambda}{\iint_{-\infty}^{\infty}\int_\Lambda PSF_\lambda(u,v)T_c(\lambda)d\lambda dudv}, c \in \{r,g,b\}, \quad (6)$$

where the monochromatic PSF_λ is averaged over λ with the weights $T_c(\lambda)$.

Thus, for PSF-based RGB imaging, we take into consideration the spectral properties of the sensor and in this way obtain accurate modeling of image formation [20]. The OTF for (6) is calculated as the Fourier transform of $PSF_c(u,v)$:

$$OTF_c(f_x,f_y) = \iint_{-\infty}^{\infty} PSF_c(u,v)e^{-j2\pi(f_x u+f_y v)}dudv, \quad (7)$$

where (f_x,f_y) are the Fourier frequency variables.

From PSFs to Imaging

Let us introduce $PSFs$ for defocus scenarios with the notation $PSF_{c,\delta}(x,y)$, where δ is a defocus distance in d_1, such that $d_1 = d_1^0 + \delta$ with d_1^0 equal to the focal distance between the aperture and the object. Introduce a set $\mathscr{D}$ of defocus values $\delta \in \mathscr{D}$

defining the area of the desirable EDoF. It is worth noting that the corresponding optical transfer functions are used with the notation $OTF_{c,\delta}(f_x, f_y)$. The definition of $OTF_{c,\delta}(f_x, f_y)$ corresponds to (7), where PSF_c is replaced by $PSF_{c,\delta}$. Thus, let $I^s_{c,\delta}(u,v)$ and $I^o_c(u,v)$ be wavefront intensities at the sensor (registered focused/misfocused images) and the intensity of the object (true image), respectively. Then, $I^s_{c,\delta}(u,v)$ are obtained by convolving the true object-image $I^o_c(u,v)$ with $PSF_{c,\delta}(u,v)$ forming the set of misfocused (blurred) color images

$$I^s_{c,\delta}(x,y) = PSF_{c,\delta}(x,y) \circledast I^o_c(x,y), \tag{8}$$

where $\circledast$ stays for convolution. In the Fourier domain, we have

$$I^s_{c,\delta}(f_x, f_y) = OTF_{c,\delta}(f_x, f_y) \cdot I^o_c(f_x, f_y). \tag{9}$$

The indexes (o,s) stay for object and sensor, respectively.

EDoF Image Reconstruction

For image reconstruction from the blurred data $\{I^{s,k}_{c,\delta}(f_x, f_y)\}$, we use a linear filter with the transfer function H_c which is the same for any defocus $\delta \in \mathscr{D}$. We formulate the design of the inverse imaging transfer function H_c as an optimization problem

$$\hat{H}_c \in \underset{H_c}{\arg\min} \underbrace{\frac{1}{\sigma^2}\sum_{\delta,k,c} \omega_\delta ||I^{o,k}_c - H_c \cdot I^{s,k}_{c,\delta}||_2^2 + \frac{1}{\gamma}\sum_c ||H_c||_2^2}_{J}, \tag{10}$$

where $k \in K$ stays for different images, $I^{o,k}_c$ and $I^{s,k}_{c,\delta}$ are sets of the true and observed blurred images (Fourier transformed), c for color, σ^2 stands for the variance of the noise, and γ is a Tikhonov regularization parameter. The parameters $\omega_\delta > 0$ are the residual weights in (10). We calculate these weights as the exponential function $\omega_\delta = exp(-\mu \cdot |\delta|)$ with the parameter $\mu > 0$. The norm $|| \cdot ||_2^2$ is Euclidean defined in the Fourier domain for complex-valued variables.

Thus, we aimed to find H_c such that the estimates $H_c \cdot I^{s,k}_{c,\delta}$ would be close to FT of the corresponding true images $I^{o,k}_c$. The second summand stays as a regularizer for H_c. Due to (9), minimization on H_c is straightforward leading to

$$\hat{H}_c(f_x, f_y) = \frac{\sum\limits_{\delta \in \mathscr{D}} \omega_\delta OTF^*_{c,\delta}(f_x, f_y)}{\sum\limits_{\delta \in \mathscr{D}} \omega_\delta |OTF_{c,\delta}(f_x, f_y)|^2 + \frac{reg}{\sum_k |I^{o,k}_c(f_x, f_y)|^2}}, \tag{11}$$

where the regularization parameter reg stays for the ratio σ^2/γ.

Therefore, the reconstructed images are calculated as

$$\hat{I}^{o,k}_c(x,y) = \mathscr{F}^{-1}\{\hat{H}_c \cdot I^{s,k}_{c,\delta}\}, \tag{12}$$

where $\mathscr{F}^{-1}$ models the inverse Fourier transform. For the exponential weight $\omega_\delta = exp(-\mu \cdot |\delta|)$, $\mu > 0$ is a parameter that is optimized. The derived OTFs (11) are optimal to make the estimates (12) efficient for all $\delta \in \mathscr{D}$, in this way, we are targeted on EDoF imaging.

MPM Modeling and Design Parameters

In our design of MPM, we follow the methodology proposed in [20]. The following parameters characterize the free-shape piece-wise invariant MPM: h is the thickness of the varying part of the mask, and N is the number of levels, which may be of different heights.

Absolute Phase Model

The proposed absolute phase $\varphi_{\lambda_0,\alpha}$ for our MPM takes the form

$$\varphi_{\lambda_0,\alpha}(x,y) = \frac{-\pi\alpha}{\lambda_0 f_{\lambda_0}}(x^2+y^2) + \beta(x^3+y^3) + \sum_{r=1, r\neq 4}^{R} \rho_r P_r(x,y). \tag{13}$$

The factor with λ_0 in this equation is introduced for the proper scaling of the MPM's quadratic phase with the phase delay of the refractive lens. The parameter α in this factor controls the optical power sharing between the lens and MPM. The cubic phase of a magnitude β is a typical component for EDoF, and the third group of the items is for parametric approximation of the free-shape MPM using the Zernike polynomials $P_r(x,y)$ with coefficients ρ_r to be estimated. We exclude from this approximation the fourth Zernike polynomial defining the quadratic defocus term because it is considered as the first item in $\varphi_{\lambda_0,\alpha}(x,y)$.

Fresnel Order (thickness of MPM)

In radians, the mask thickness is defined as $Q = 2\pi m_Q$, where m_Q is called 'Fresnel order' of the mask which in general is not necessarily integer. The phase mask profile of the thickness Q is calculated as

$$\hat{\varphi}_{\lambda_0,\alpha}(x,y) = mod(\varphi_{\lambda_0,\alpha}(x,y) + Q/2, Q) - Q/2. \tag{14}$$

The operation in (14) returns $\hat{\varphi}_{\lambda_0,\alpha}(x,y)$ taking the values in the interval $[-Q/2,\ Q/2)$. The parameter m_Q is known as 'Fresnel order' of the mask. For $m_Q = 1$, this restriction to the interval $[-\pi,\ \pi)$ corresponds to the standard phase wrapping operation.

Number of Levels

The mask is defined on $2D$ grid (X,Y) with the computational sampling period (computational pixel) Δ_{comp}. We obtain a piece-wise invariant surface for MPM after the non-linear transformation of the absolute phase. The uniform grid discretization of the wrapped phase profile $\hat{\varphi}_{\lambda_0,\alpha}(x,y)$ to the N levels is performed as

$$\theta_{\lambda_0,\alpha}(x,y) = \lfloor \hat{\varphi}_{\lambda_0,\alpha}(x,y)/N \rfloor \cdot N, \tag{15}$$

where $\lfloor w \rfloor$ stays for the integer part of w. The values of $\theta_{\lambda_0,\alpha}(x,y)$ are restricted to the interval $[-Q/2,\ Q/2)$. Q is an upper bound for thickness phase of $\theta_{\lambda_0,\alpha}(x,y)$.

The introduced discretization and modulo functions are not differentiable, therefore we use a smoothing approximation to be able of optimizing the thickness and the number of levels of MPM by gradient descent algorithms. The details of this approximated function can be found in [16].

The mask is designed for the wavelength λ_0. Thus, the piece-wise phase profile of MPM for the wavelength λ is calculated as

$$\varphi_{MPM_{\lambda_0,\lambda,\alpha}}(x,y) = \frac{\lambda_0(n(\lambda)-1)}{\lambda(n(\lambda_o)-1)}\theta_{\lambda_0,\alpha}(x,y), \tag{16}$$

where $\theta_{\lambda_0,\alpha}$ is the phase shift of the designed MPM and $n(\lambda)$ is the refractive index of the MPM material, $x \in X, y \in Y$. The MPM thickness h in length units is of the form

$$h_{\lambda_0}(x,y) = \frac{\lambda_0}{(n(\lambda_o)-1)}\frac{\theta_{\lambda_0,\alpha}}{2\pi}. \tag{17}$$

Optimization Framework

Figure 2 illustrates the framework for optimizing the proposed optical system using iterative NN algorithms and stochastic gradient ADAM optimization. It can be downloaded from PyTorch with an optimized tensor library for Neural Network (NN) learning using GPUs [1]. Some details concerning this framework are given in what follows in this section.

Loss Function

Let Θ be a full set of the optimization parameters defined as

$$\Theta = (\alpha, \beta, \rho_r, reg). \tag{18}$$

Then, we use the following multi-objective formulation of our optimization goals

$$\hat{\Theta} = \arg\max_{\Theta}(PSNR(\Theta,\delta), \delta \in \mathscr{D}). \tag{19}$$

In this formulation, we maximize all $PSNR(\Theta,\delta)$, $\delta \in \mathscr{D}$, simultaneously, i.e. to achieve the best accuracy for all focus and defocus situations. Here, $PSNR(\Theta,\delta)$ is calculated as the mean value of $PSNR^k(\Theta,\delta)$ over the set of the test-images, $k \in K$:

$$PSNR(\Theta,\delta) = mean_{k\in K}(PSNR^k(\Theta,\delta)). \tag{20}$$

There are various formalized scalarization techniques reducing the multi-objective (vector) criterion to a scalar one. Usually, it is achieved by aggregation of multiple criteria in a single one (e.g. [21]). In this paper, we follow pragmatical heuristics comparing $PSNR(\hat{\Theta},\delta)$ as the $1D$ functions of δ in order to maximize $PSNR(\Theta,\delta)$ for each $\delta \in \mathscr{D}$. Here, $\hat{\Theta}$ are estimates of the optimization parameter. In this heuristic, we follow the aim of the multi-objective optimization (19). In developing the proposed optimization framework, the main challenges included satisfying manufacturing constraints, finding stable optimization algorithms, and fitting models within memory constraints.

Parameters for MPM design and simulation tests

The sensor's parameters used in simulation correspond to the physical sensor used in our experiments: pixel size 3.45 μm and resolution 512×512 pixels. The Fourier transform for PSFs calculations are produced on the grid 3000×3000 of the computational pixel size Δ_{comp}=2 μm, defining discretization of lens and MPM. We fixed the number of MPM levels to $N = 52$ and Fresnel order to $m_Q = 1$, the latter restricts the MPM phase wrapping to the interval $[-\pi, \pi)$. The optimization stage includes finding the optimal α, β, ρ_r for the MPM design and reg for the image inverse reconstruction using the Adam stochastic gradient descent solver with the step-size 5×10^{-3}.

We analyze and compare the hybrid optics of different lens diameters (aperture size of hybrid) taking values $(5, 6, 7, 9)$ mm and lens focal length taking values $f = (3, 5, 7, 10)$ mm. The focus imaging distance for the hybrid is fixed to $d_1^0 = 1$ m. For each lens focal length f, d_2 is calculated according to the focusing equation $\frac{1}{d_1^0} + \frac{1}{d_2} = \frac{1}{f}$. These values of d_2 are very close to f. It was concluded from our tests that $R = 14$ (Zernike coefficients excluding the fourth polynomial in (13)) is enough and larger values of R do not improve image quality significantly. The design wavelength is $\lambda_0 = 510$ nm. An additive white Gaussian noise is included in observations with variance equal to 1×10^{-4}. We choose 31 wavelengths, with step 10 nm, covering the visual interval $(400-700)$ nm to model RGB imaging. To enable EDoF imaging, we use Wiener filtering with $d_1 = 0.5, 0.6, 0.7, 1.0, 1.9, 10,$ and 100.0 m. These d_1 define the defocus parameter δ in (11) as $\delta = d_1 - d_1^0$. The optimization stage employs 200 epochs, which takes approximately 6 hours on NVIDIA GeForce RTX 3090 GPU with a memory of 24GB.

Data sets for optimization and tests

For optimization and training, we chose 3550 high-resolution RGB images from databases [2]. For testing the designed systems, we used 200 high-resolution RGB images from the same databases which are not included in the training set. In what follows, all illustrative materials (tables, curves, and images) are given for these test images.

Simulation Tests and First Stage of Optimization

In this section, we design the phase profiles for MPM in the hybrid optical setup with different aperture sizes (5, 6, 7, and 9) mm and lens focal lengths (3. 5, 7, and 10) mm. Our intention is to find combinations of these physical parameters for the best achromatic EDoF imaging. The corresponding numerical results obtained by simulation using the end-to-end joint optimization of optics and inverse imaging algorithms are presented in Table 1. The reported $PSNRs$ are averaged over 7 depth (defocus) distances d_1 from the interval (0.5 - 100.0) m and over 200 RGB test images.

The imaging accuracy is evaluated and reported in two versions: $PSNR_{RGB}$ calculated for each of the color channels separately (column 4), and $PSNR_{total}$ calculated for all three color channels jointly (column 3). The best result (highest values of PSNR) is achieved by the setup with 6 mm aperture size and 5 mm lens focal length. These physical parameters result in F-number=0.83 and a 70.5-degree field of view (FOV). The $PSNR_{total}$ value for this case is equal to 44.23 dB, but it degrades dramatically for larger and smaller focal lengths within the fixed diameter. If we compare the PSNR for the color channels separately, the values for 6mm diameter designed hybrid optics are highest (all above 41 dB) and more or less the same for all color channels.

[1] The Pytorch library https://pytorch.org/

[2] https://data.vision.ee.ethz.ch/cvl/DIV2K/, and http://cv.snu.ac.kr/research/EDSR/Flickr2K.tar.

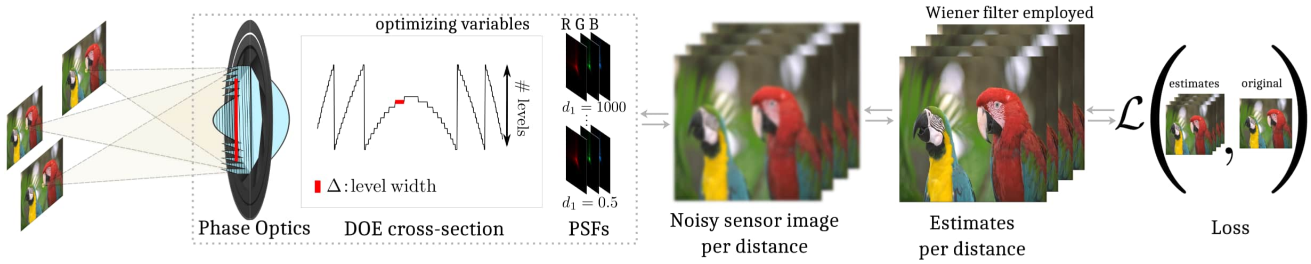

Figure 2. The optimal design framework of phase-encoded optics and image reconstruction algorithms for achromatic EDoF. The spectral PSFs are convolved with batches of RGB ground-truth images. Inverse imaging provides estimates of these images. Finally, a quality/accuracy loss $\mathscr{L}$, such as mean squared error with respect to the ground-truth images (or PSNR criterion), is defined on reconstructed images.

Comparative performance of the hybrid optics: different lens diameter (aperture size) and lens focal length.

Diameter (mm)	Focal length (mm)	$PSNR_{total}$(dB)	PSNR per channel			F-number	FOV (degree)
			R	G	B		
5	3	31.48	28.43	34.61	31.64	0.6	99.3
	5	**41.58**	43.20	44.65	39.82	1	70.5
	7	38.75	39.44	42.71	35.98	1.4	53.6
	10	36.29	36.98	40.11	31.84	2	38.9
6	3	25.64	23.12	27.21	22.89	0.5	99.3
	5	**44.23**	**44.92**	**46.81**	**41.74**	**0.83**	**70.5**
	7	36.61	39.41	40.87	30.29	1.17	53.6
	10	33.66	32.22	34.07	29.46	1.66	38.9
7	3	25.8	25.29	29.77	21.58	0.43	99.3
	5	33.28	29.09	37.09	29.47	0.71	70.5
	7	**36.14**	34.61	39.26	29.93	1	53.6
	10	31.65	28.41	33.20	29.86	1.43	38.9
9	3	24.21	24.74	27.46	19.59	0.33	99.3
	5	26.43	22.45	28.40	26.05	0.54	70.5
	7	**34.79**	29.08	39.10	30.19	0.76	53.6
	10	30.97	31.00	36.25	26.09	1.08	38.9

Note also, that for each lens diameter there is an optimal lens focal length and this optimal value is close to the diameter size. The optimal focal lengths for the diameters (5, 6, 7, and 9) mm are (5, 5, 7, and 7) mm, respectively. We may conclude that the lens focal length plays a crucial role in hybrid optics and there is a trade-off between imaging quality and FOV. Smaller focal length (in Table 1, 3 mm) gives wider FOV at expense of less imaging accuracy. This conclusion is valid for all lens diameters in Table 1.

Further information on the comparative performance of the imaging system with the optimized hybrid optics can be seen in Figure 3. Here we present PSNR curves as functions of d_1 (distance between the object and optics) averaged over 200 test images. The four curves are given for the four values of lens diameter with the corresponding optimal lens focal length as shown in Table 1.

The uniformly best performance is achieved by the 6 mm aperture hybrid optics with $f_0 = 5mm$. For this case, the PSNR value is about 37dB for the defocus point $d_1 = 0.5m$. The peak of this curve is at $d_1 = 1.0m$ with PSNR=50dB. Remind, that this is a focus point of the system. For larger defocus distances, $d_1 > 1$, PSNR takes lower values which are nevertheless close to 45 dB, which guarantees high-quality imaging. The hybrid with the 5 mm aperture and $f_0 = 5mm$ also demonstrates a very good performance with slightly lower PSNR values. For the two other cases: $D = 7, f_0 = 7$ mm and $D = 9, f_0 = 7$ mm, we can see a much worse performance with PSNR values lower from 5 to 10 dB as compared with the best ones.

The spectral performance of the best-optimized hybrid system ($D = 6mm$ and $f_0 = 5mm$) characterized by PSNRs calculated for the RGB channels as functions of d_1 is presented in Figure 4. These curves with $PSNRs$ averaged over 200 test-images show the accuracy of imaging for each color channel and depth d_1. The $PSNR_{total}$, black curve in Figure 4, shows the accuracy as a function of d_1 calculated for all spectral channels simultaneously as averaged over 200 test-images. The color channel curves mainly follow the behavior of $PSNR_{total}$. All these spectral curves are well above the 35 dB line confirming high-accuracy imaging for all d_1 and all spectral channels.

Figure 5 illustrates a visual performance of the designed hybrid systems of different diameters ($D = 5, 6, 7, 9$) mm with the optimal lens focal length as defined in Table 1. The reconstructed images and their small fragments are shown for the distances $d_1 = (0.5, 1.0, 100.0)$ m. The color channel's PSNR values are shown in these images. Thus, the comparison can be produced visually and numerically. The optimized phase profiles of MPMs are shown in this first row of Figure 5.

Comparing these results, we may conclude, that the best re-

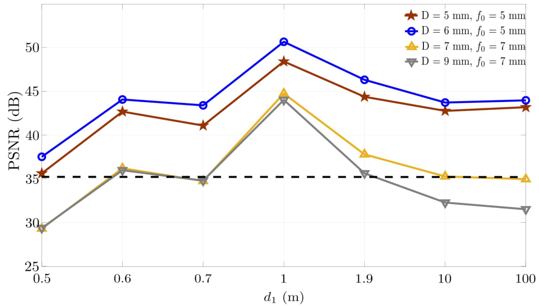

Figure 3. PSNR curves of the optimized hybrid setups with 4 different aperture size D= (5, 6, 7, and 9) mm as a function of distance from the scene to the optics (d_1). The optimized hybrid setups with 5 and 6 mm diameters perform in the best way with more or less uniform PSNR values which are well above the good imaging quality line, PSNR = 35 dB, for all depths. The advantage of hybrid optics with $D = 6$ mm versus $D = 5$ mm is obvious of about 1 to 2 dB of PSNR values for each distance. The imaging with $D = (7$ and $9)$ mm shows good results in the vicinity of the system focal point ($d_1 = 1m$), but the performance is dropped for far and even quite close distances.

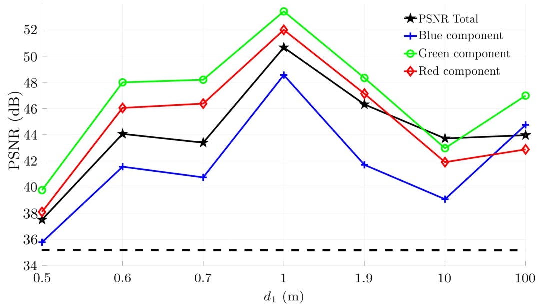

Figure 4. The spectral performance of the best-optimized hybrid system ($D = 6mm$ and $f_0 = 5mm$) is characterized by PSNRs calculated for the RGB channels as functions of d_1. All curves are above the good imaging quality line of 35 dB. The curves for color components mainly follow the behavior of the total PSNR curve (black).

sults are achieved by the 5 mm and 6 mm diameter aperture sizes (columns 2 and 3) with an advantage of the latter one. For instance, for $d_1 = 0.5m$, the improvement in PSNR is about 2 to 4 dB for color channels in favor of the hybrid optics with a 6 mm lens diameter. Moreover, details and colors are better preserved in this case. This best setup provides uniformly better imaging quality for various depths and colors. The zoomed fragments of the reconstructed images visually reveal clearly that the hybrid optics with 7 and 9 mm diameters (columns 4 and 5) are suffering from strong chromatic aberrations and are quite blurry.

The advantage of the best hybrid optics with $D = 6$ mm and $f_0 = 5$ mm is well seen as compared with its counterparts, which is in direct agreement with the results shown in Table 1. Additionally in Figure 6, for this best hybrid system, we show the cross-sections of PSFs for the three RGB channels and for the distances d_1 used in Figure 5. These cross-section curves are well consolidated, which explains a source of the good performance of the imaging system for different distances d_1 and different color channels. The forthcoming optimization results and experimental tests in the following sections are produced for the hybrid optics with the found optimal $D = 6$ mm, $f_0 = 5$ mm.

Physical Experiments and Second/Third Stages of Optimization

Optical Setup and Equipment

In this work, to implement our hybrid optics and in order to avoid building several MPM to physically analyze the performance of our camera, we build an optical setup based on a programmable phase SLM to exploit its phase capabilities to investigate the performance of the designed hybrid setup. The optical setup is depicted in Figure 7(a), where 'Scene' denotes objects under investigation; the polarizer, 'P', keeps the light polarization needed for a proper wavefront modulation by SLM; the beam-splitter, 'BS', governs SLM illumination and further light passing; the lenses 'L_1' and 'L_2' form a 4f-telescopic system transferring the light wavefront modified by SLM to the lenses 'L_3' and 'L_4' plane; the lenses 'L_3' and 'L_4' forms an image of the 'scene' on the imaging detector, 'CMOS'. We use two lenses 'L_3' and 'L_4' tightly fixed to each other in order to get the hybrid's lens, as in Figure 1, of a smaller focal length: $f_0 = f_1/2$, where f_1 is the focal length of 'L_3' and 'L_4'.

For physical modeling of MPM phase delay, we use SLM: the Holoeye phase-only GAEA-2-vis SLM panel, resolution 4160×2464, pixel size 3.74 μm. 'L_1' and 'L_2' are achromatic doublet lenses with diameter 12.7 mm and focal length 50 mm; Two BK7 glass lenses 'L_3' and 'L_4'are of diameter 6 mm and focal length 10.0 mm which results in $f_0 = 5.0$ mm; 'CMOS' Blackfly S board Level camera with the color pixel matrix Sony IMX264, 3.45 μm pixel size and 2448×2048 pixels. This SLM allows us to experimentally study optical hybrid imaging with an arbitrary phase-delay distribution for the designed MPM. The MPM 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 2.0π for wavelength equal to 510 nm. This 2.0π corresponds to the value 255 of **.bmp* file for the phase-delay image of MPM.

Figure 7(a) illustrates the architecture of the developed optical setup with SLM and the photo of the corresponding hardware. Figure 7(b) shows the photo of this hardware with three monitors displaying the scenes (images) with three fixed distances $d_1 = (0.5, 1.0, 1.8)$ m. The imaging monitors have a resolution of 1920×1080 and 570ppi. The distance $d_1 = 1.0$ m is the focal point of the optical system.

Optimization of image reconstruction provided a fixed MPM: learning-based approach (Second stage of optimization)

This optimization is used in our physical experimental works provided that the optimized phase-delay profile of MPM with $D = 6$ mm, obtained in the model-based approach, is implemented by SLM. The outputs of the sensor are blurred images registered for a sequence of the train dataset images displayed on three monitors at three different depths, $d_1 = (0.5, 1.0, 1.8$ m). Convolutional Neural Network (CNN) is used to fit these blurred images to the known true target images. In this way, CNN designs the inverse imaging algorithm defined by the CNN parameters. For optimization, we exploit the stochastic gradient ADAM optimizer. The training process is running for 3550 high-quality images on three

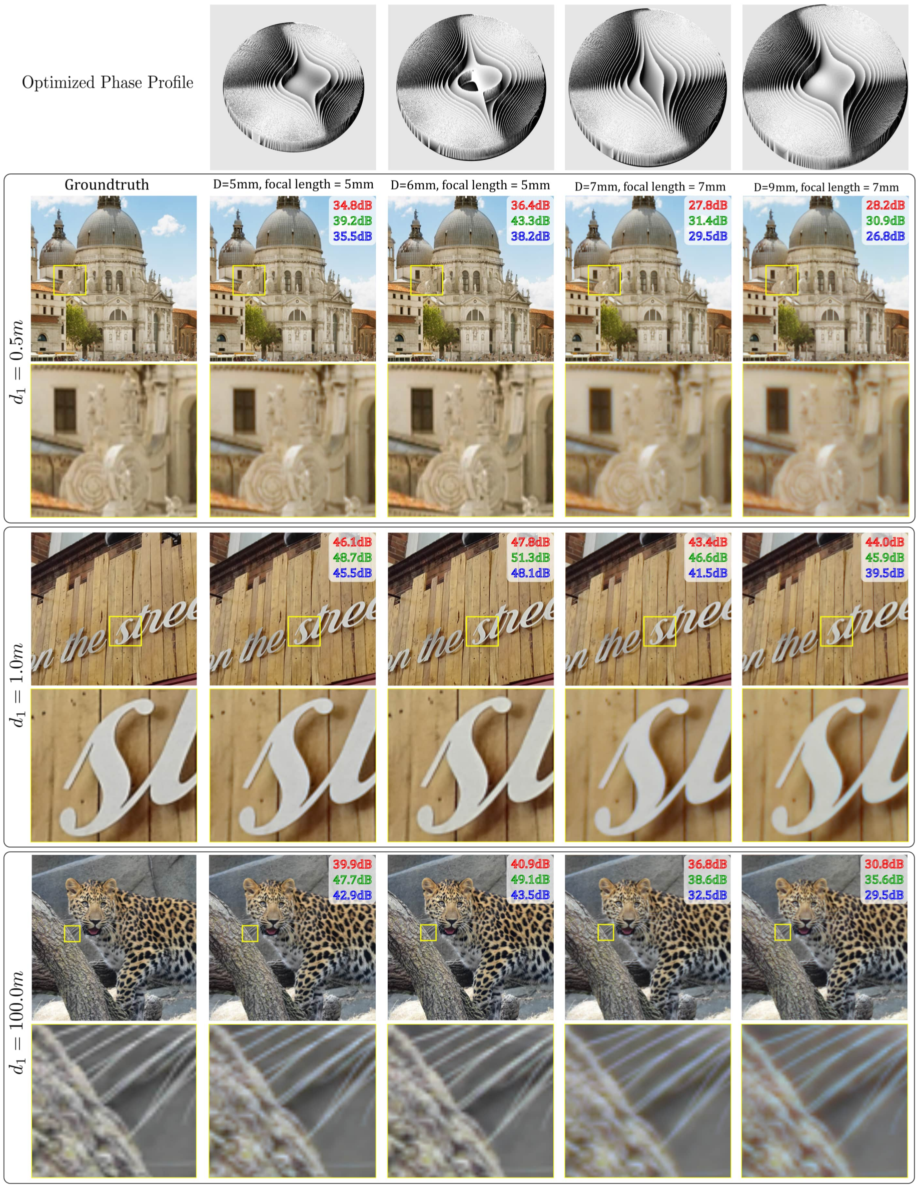

Figure 5. *Visual performance of the designed hybrid systems is illustrated for different diameters ($D = 5,6,7,9$) mm with the optimal lens focal length as defined in Table 1. The reconstructed images and their small fragments are shown for the distances $d_1 = (0.5, 1.0, 100.0)$ m. The color channel's PSNR values are shown in these images. Thus, the comparison can be produced visually and numerically. High-quality imaging for different colors and depths is achieved by the optical hybrid setups with 6 and 5 mm diameter and 5 mm focal lengths (columns 2 and 3). In contrast, the results for 7 and 9 mm diameters (columns 4 and 5) are suffering from strong chromatic aberration and the performance is degrading especially for off-focus distances $d_1 = 0.5$ and $d_1 = 100.0$ m. The optimized phase profiles of MPMs are shown in the first row of the image.*

monitors which gives in a total of 10650 registered images. The network has been trained for 320 epochs which takes two weeks on NVIDIA GeForce RTX 3090 GPU.

Figure 8 illustrates an architecture of CNN used in our experiments (DRUNet CNN [22]). 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) downscaling and 2×2 transposed convolution (TConv) upscaling 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 downscaling and upscaling of each scale. Each residual block only contains one ReLU activation function. The proposed DRUNet is bias-free, which means no bias is used in all the Conv, SConv and TConv layers [22].

An appropriate loss function is required to optimize the inverse imaging to provide the desired output. Thus, we use a weighted combination of PSNR between estimated and ground truth images ($\mathscr{L}_{PSNR}$), perceptual loss, and adversarial loss which

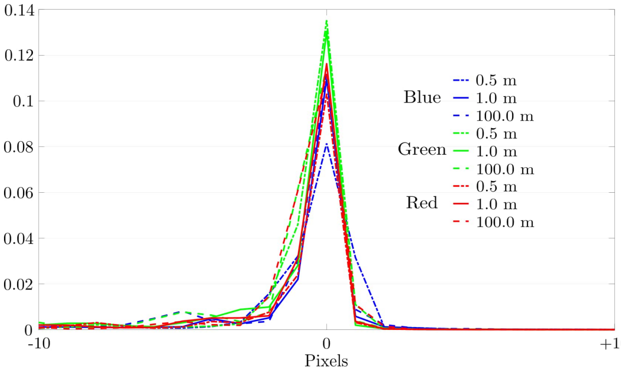

Figure 6. *For the best hybrid system ($D = 6$ mm, $f_0 = 5$ mm), we show the cross-sections of the spectral PSFs for the three RGB channels and for the distances d_1 used in Figure 5. These cross-section curves are well consolidated, which explains the good performance of the imaging system for different distances d_1 and different color channels.*

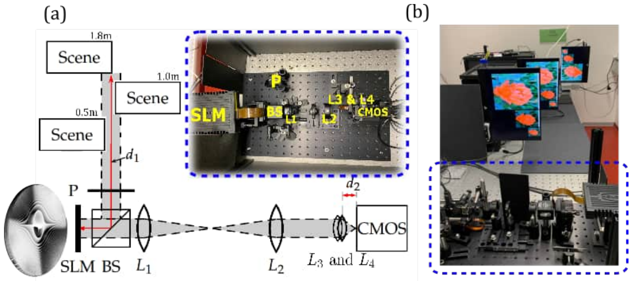

Figure 7. *Experimental optical setup. Figure (a) illustrates the architecture of the optical setup with SLM and the photo of the corresponding hardware. P is a polarizer, BS is a beamsplitter, SLM is a spatial light modulator. The lenses L_1 and L_2 form the $4f$-telescopic system projecting wavefront from the SLM plane to the imaging lenses L_3 and L_4, CMOS is a registering camera. d_1 is the distance between the scene and the plane of the hybrid optics (L_3 and L_4) and d_2 is the distance between this hybrid optics and the sensor. Figure (b) shows the photo of this hardware with three monitors displaying the scenes (images) of three fixed distances $d_1 = (0.5, 1.0.1.8)$ m.*

are given below.

Perceptual loss: To measure the semantic difference between the estimated output and the ground truth, we use a pretrained VGG-16 [23] model for our perceptual loss [24]. We extract feature maps between the second convolution (after activation) and second max pool layers φ_{22}, and between the third convolution (after activation) and the fourth max pool layers φ_{43}. Then, the loss $\mathscr{L}_{Percep}$ is the averaged PSNR between the outputs of these two activation functions for both estimated and ground truth images.

Adversarial loss: Adversarial loss [25] was added to further bring the distribution of the reconstructed output close to those of the real images. Given the swish activation function [26] as our discriminator D, this loss is given as $\mathscr{L}_{Adv} = -\log(D(I_{est}))$ where I_{est} models the estimated image.

Our total loss for the proposed CNN inverse imaging while training is a weighted combination of these three losses and is given as, $\mathscr{L}_{CNN} = \sigma_1 \mathscr{L}_{PSNR} + \sigma_2 \mathscr{L}_{Percep} + \sigma_3 \mathscr{L}_{Adv}$, where, σ_1, σ_2 and σ_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 are to be optimized.

In Figure 9 we report an evaluation of PSNR versus a number of epochs. From these results, we can see that the quality achieved by CNN for the designed hybrid system is quite high for the training data set. Illustrating reconstructed images chosen among the testing dataset are presented for epochs 0, 100, and 320. It could be seen that the trained network performs well and the output image for epoch 320 is sharp enough. The practical value of this approach to image processing design follows from using physical modeling of image formation including in particular wavefront propagation and mosaicing/demosaicing operations.

Experimental Results: model-based SLM

In this section, we present the results of two types of experiments. In the first one, the test-images are displayed on the three monitors as in Figure 7(b), the observations are blurred and the images are reconstructed by the trained CNN. These results are shown in Figure 10. In this scenario, we presented and evaluated the quality of reconstructions visually as well as numerically by PSNR values for each of the RGB color channels.

In the second type of experiments, we image a scene composed of different objects arbitrarily located within the range (0.4-1.9) m from the hybrid optics. This optical setup is used to evaluate the performance of the designed system in a real-world scenario for the EDoF imaging task. The performance of the designed system is compared with the compound multi-lens commercial smartphone camera. These results can be seen in Figure 11.

The results in Figure 10 are presented in 7 columns for three depth distances: $d_1 = 0.5\ m$ (columns 2 and 3), $d_1 = 1.0\ m$ (columns 4 and 5), and $d_1 = 1.8\ m$ (columns 6 and 7). The Groundtruth column shows the true images. Two images from the test dataset are presented in this figure for comparison (rows 1 and 3) with one zoomed region (rows 2 and 4). We can see the zoomed fragments of the blurred noisy images on the sensor used for CNN image reconstruction as well as the corresponding reconstructed images. The zoomed sections for blurry and reconstructed images visually reveal that the images are sharp and clear enough and the quality of imaging is high and more or less the same for different depths. Besides, the colors are well preserved properly along with distances. If we compare the results numerically by PSNR, we could conclude that the PSNR values for different colors and depths are more or less the same at about 23 dB. It confirms that the designed hybrid imaging indeed demonstrates achromatic EDoF imaging.

The imaging results for the real-scene scenario are presented in Figure 11. The scene consists of 5 objects located at different distances from 0.4 m to 1.9 m, approximately: $d_1 = 0.4\ m$ (Train Wagon), 0.65 m (Locomotive), 1.15 m (ThorLab snack box), 1.2 m (Dwarf Christmas Santa Claus Doll), and 1.9 m (Panda toy). It is worth mentioning, that for smartphone (compound optics), we adjusted the focusing distance to $d_1 = 1.0$ m as it is for the hybrid system. The hybrid diffractive imaging is compared with imaging by the smartphone camera (column 1).

For the designed hybrid, two image reconstruction techniques are demonstrated: model-based (column 2) and learning-based (column 3). In the model-based algorithm, the scene is reconstructed using the calculated color channel PSFs and the inverse imaging according to Eq. (12). After this step, a denoising

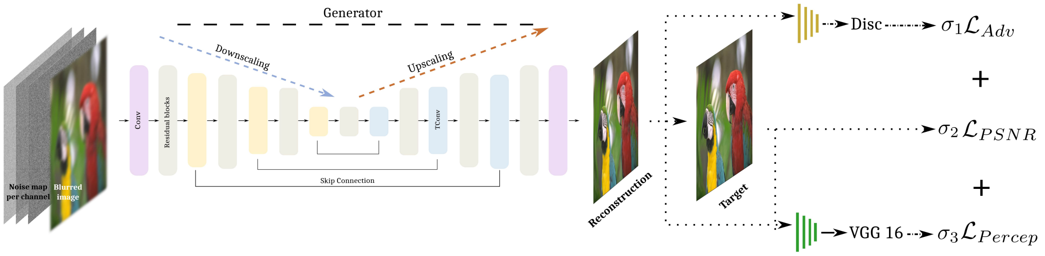

Figure 8. *Inverse imaging UNet-based neural network architecture. The generator model is a U-net architecture that has seven scales with six consecutive downsampling and upsampling operations* [22]*. We adopt a weighted combination of PSNR between estimated and ground truth images, $\mathscr{L}_{PSNR}$, and perceptual losses $\mathscr{L}_{Adv}$ and $\mathscr{L}_{Percep}$, with weights σ_1, σ_2, and σ_3.*

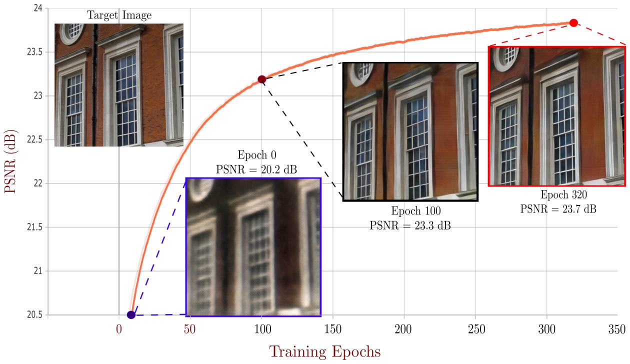

Figure 9. *Performance of CNN for design of inverse imaging algorithm. The quality achieved by CNN starts from 20.5 dB and reaches 23.8 dB of PSNR for the training image set. The reconstructed images over the testing dataset are presented for three epochs 0, 100, and 320 for visualization of the training process. It could be seen that the trained network performs well for this task and the output image for epoch 320 is sharp enough.*

process equipped with a sharpening procedure [27] is performed over the estimated scene to improve the quality of imaging. This final denoised image is returned as the estimated scene from experimental data. Contrary to it, the learning-based inverse imaging uses the trained UNet. For a detailed comparison, the four zoomed fragments of the images are shown in rows 4, 5, 6, and 7 which correspond to the scene's objects of different out-of-focus distances. Comparing columns 2 and 3, we may note an obvious advantage of learning-based inverse imaging. The model-based approach is not able to recover all details, the output image is still blurry, and chromatic aberrations are strong. There are a number of reasons for this huge gap. First of all, it concerns a mismatch between reality and the analytical modeling of image formation by PSFs. Second, the mosaicing/demosaicing are not included in our modeling. The leaning-based approach allows successfully compensate these drawbacks of the analytical modeling.

Comparison of the learning-based hybrid imaging (column 3) versus the smartphone camera imaging (column 1) results in an exciting conclusion about a quite clear advantage of hybrid diffractive imaging. This advantage is obvious in the sharpness of images for all distances. Thus, hybrid imaging demonstrates high-quality all-in-focus imaging. Concerning color aberrations: red and green perhaps not be properly presented by the hybrid but the white color is definitely perfect. Thus overall, hybrid diffractive imaging can be tread at least as quite competitive and even advanced with respect to the commercial smartphone with multi-lens optics. Here we need to note that the white balance and γ correction procedure have been produced for the images reconstructed by the hybrid system in order to have a fair comparison with the smartphone camera.

Experimental Results: optimized HIL-SLM (Third stage of optimization)

In this section, we use the HIL methodology for the joint design of optimal phase pattern for SLM and image reconstruction algorithm. Both DOE design and NN techniques for imaging are already used jointly in the frame of HIL methodology in our recent work [17] for imaging systems with an aperture of 9.2 mm and the lens focal distance 10 mm. The experimental results shown in [17] demonstrated a serious advantage in imaging numerically and visually versus the model-based design as well as compared with the compound commercial multi-lens optics in all-in-focus imaging for the depth range of 0.4-1.9 m.

In this section, we produce the third stage optimization of SLM profile and image reconstruction algorithm for 6 mm aperture and 5 mm focal length. Following the mainstream of the approach proposed in [17], we use the phase profile designed by the model-based end-to-end optimization approach (optimization stage 1) as the starting point of this third stage optimization. The initialization for the image processing is done using the algorithm also obtained at the first stage of optimization.

Optimization of SLM requires specific instruments as in the HIL setup SLM is considered as a part of the black box of an unknown mathematical model. Thus, the derivatives usually required in NN optimization cannot be calculated. For optimization of SLM, as in [17], we use the derivative-free $0th$-order stochastic evolutionary search method CMA-ES [28] which updates the shape of the phase-profile implemented on SLM (tuning α, β, and ρ_r in (13)).

It is important to emphasize that the use of a programmable phase SLM as DOE and its end-to-end optimization in the HIL-SLM setup will ensure phase modulation is properly physically modeled without any discrepancy between mathematical models and the physical reality typical for conventional model-based approaches. Overall, the design algorithm proposed in our recent work [17] follows an alternating iteration methodology: fixing the hyperparameter of SLM we produce optimization of the inverse imaging algorithm, what follows by optimization of SLM assuming that SLM parameters and so forth. The third stage of optimization in this paper assumes only two steps of this gen-

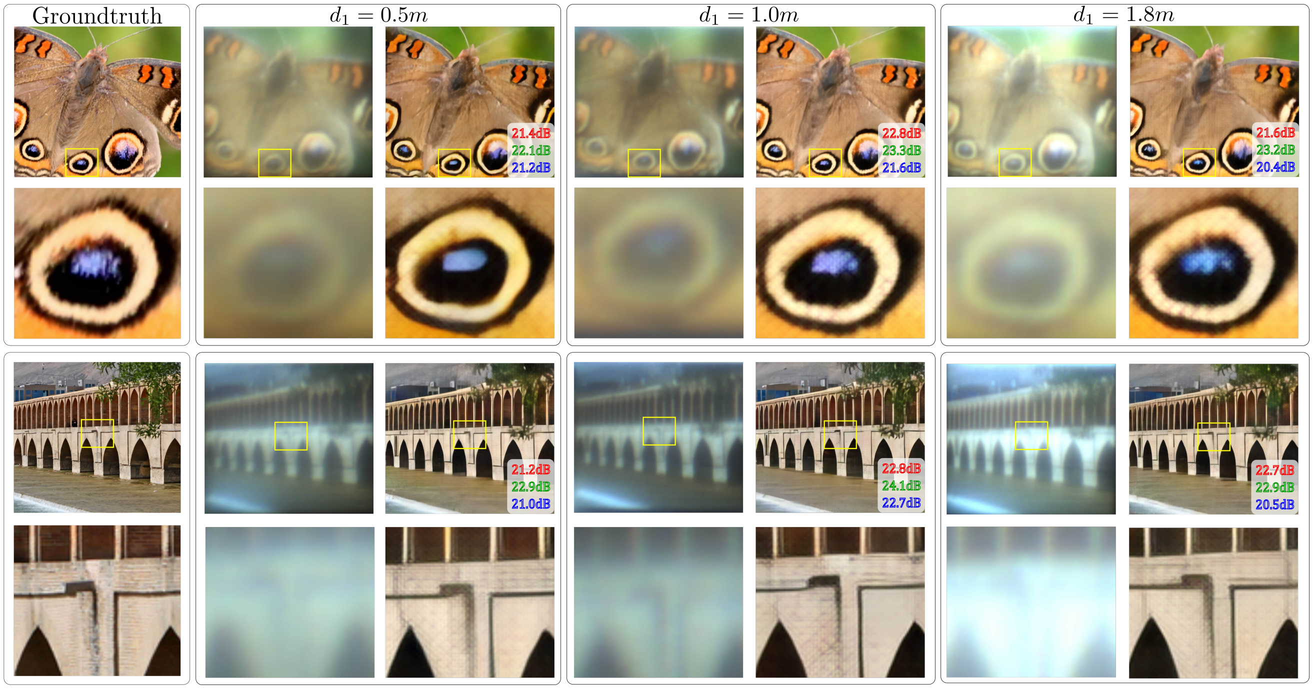

Figure 10. *Results for three monitor setup over the testing dataset. The reconstructed images with the zoomed region at three different distances (imaging monitor-SLM):* $d_1 = 0.5, 1.0, 1.8\ m$*, for the optimized hybrid system. The PSNR values are reported for each depth and each color channel separately. High-quality imaging with PSNR values of about 23 dB for different imaging depths and colors is achieved by the designed hybrid.*

eral procedure. First, we produce optimization of SLM provided the fixed image processing as found at the second stage optimization and after that, we optimize the inverse imaging assuming that SLM is fixed.

The results of this design and their comparison with those obtained provided SLM phase obtained by the model-based methodology (second stage of optimization) are shown in Figure 12 for the three monitor scenario. We wish to note that this figure is an analog to Figure 10 with results of the first and second stage optimization. The images displayed on three monitors and captured by two different phase profiles SLM (model-based and HIL-based) are chosen among the set of the test-images different from those in Figure 10. Column 1 shows the true images as displayed by the three monitors for different depths. We also presented blurred images on the sensor (columns 2 and 4) and images after reconstruction (columns 3 and 5) for each optimized system. Columns 2 and 3 correspond to model-based SLM (second stage optimization) and columns 4 and 5 correspond to the SLM optimized in the HIL methodology following optimization of image processing (third stage optimization). The rows in Figure 12 show images and their enlarged fragments obtained for different distances d_1. As can be seen, the HIL optimized phase profile of SLM is quite different from that obtained due to the model-based technique (row 1). The numerical advantage of the HIL optimized SLM follows from a comparison of PSNR values calculated for each of the RGB channels separately. The visual comparison is also in favor of the HIL optimized SLM which is in particular clear from the visualization of the image fragments.

For the real-scene experiments scenario, we arranged a scene composed of toys (different from those in Figure 11) located within the range (0.3-2.0) m from the optics. The imaging performance of the two designed diffractive imaging systems are compared also with imaging by the compound multi-lens commercial camera of smartphone. These results are depicted in Figure 13. The scene consists of 7 toys in different distances from the optics, approximately: $d_1 = 0.3\ m$ (Pine Tree), $d_1 = 0.6\ m$ (Locomotive), $d_1 = 0.8\ m$ (Mouse), $d_1 = 1.0\ m$ (Stacking Cups), $d_1 = 1.2\ m$ (Snake), $d_1 = 1.9\ m$ (Bear), and $d_1 = 2.0\ m$ (Dwarf Christmas Santa Claus Doll).

For the two hybrid designs, registered blurry images on the sensor are presented (columns 2 and 4) along with the corresponding reconstructed images (columns 3 and 5). The images obtained by the smartphone camera are shown in column 1. For more details, the four zoomed fragments of the images are shown in rows 3, 4, 5, and 6 corresponding to the scene's objects of different out-of-focus distances. Comparing the output images after reconstruction for the hybrid designs, columns 3 and 5, we may note an obvious advantage of the HIL-based design as compared with the model-based one. The color is better preserved and the images are sharper for the HIL-based design for both registered and recovered images. In comparison with smartphone camera important to note that the diffractive imaging is at least not worse. Even more, comparing the performance of the HIL-based hybrid design (column 5) versus the smartphone (column 1) we may conclude that performance of the HIL-SLM system is more or less the same for all distances d_1, while it is not true for the smartphone camera, which, in particular, shows the worse result for very close and far distances ($d_1 = 0.3\ m$, $d_1 = 1.9\ m$, and $d_1 = 2.0\ m$). On the other hand, imaging of the objects at close to the focal distance $d_1 = 1.0\ m$ (Stacking Cups, Mouse, and Snake) are sharper in the

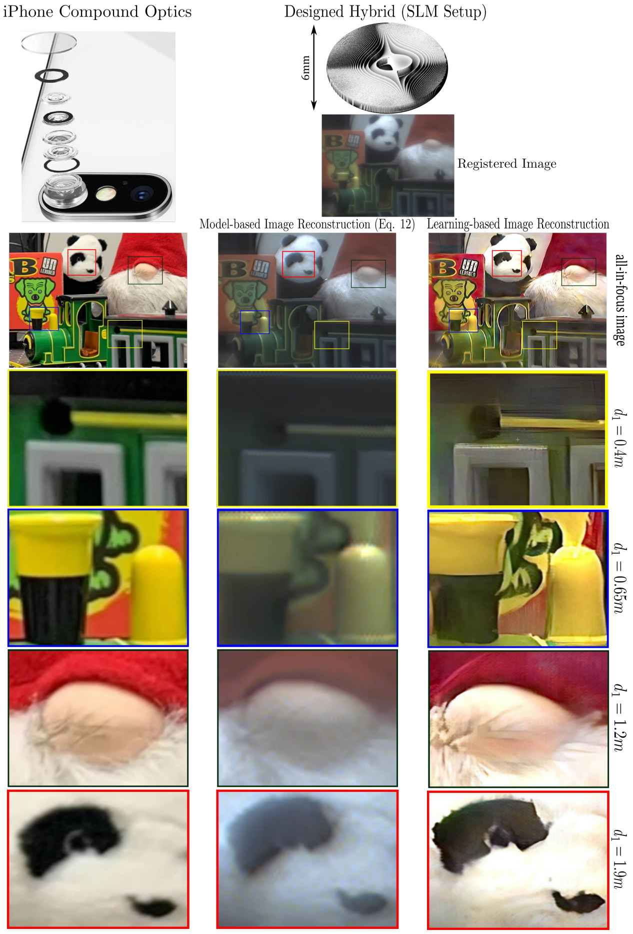

Figure 11. *Comparison of the designed hybrid diffractive imaging versus the compound lens camera of smartphone. For the designed hybrid, two image reconstruction approaches are employed (columns 2 and 3) to recover the blurred image on the sensor: Model-based and learning-based. The obtained images are presented in row 3 with their enlarged fragments in rows 4, 5, 6, and 7 corresponding to four off-focus distances $d_1 = 0.4, 0.65, 1.2, 1.9$ m, respectively. By comparing the results over the recovering approaches in the designed hybrid (columns 2 and 3), the advantage of using a deep UNet-style CNN is clear. For the smartphone camera, the imaging quality is not good for close and far distances. The visual advantage in sharpness and color preservation is clearly in favor of the designed hybrid imaging.*

smartphone image.

Conclusion

For the first time, aperture size, lens focal length, and distance between MPM and sensor are considered as optimization variables for diffractive achromatic EDoF imaging. This paper demonstrated the success of the design and end-to-end optimization proposed for computational imaging with optics composed of a single refractive lens and a diffractive phase-encoded MPM. Specifically, the designed imaging system is clearly superior to the multi-lens smartphone camera in terms of comparisons with imaging powered by the phone. An important novelty proposed in this paper is a physical modeling of MPM by SLM that allows for the online design of free-shape phase encoding for diffractive optics, thereby eliminating the mismatch between theoretical image formation models and physical reality. Experimental works confirm a strong advantage of DOE design using SLM for the programmable phase tuning in the proposed HIL methodology. Comparing with the commercial phone camera, the proposed hybrid system demonstrates quite competitive high-quality imaging.

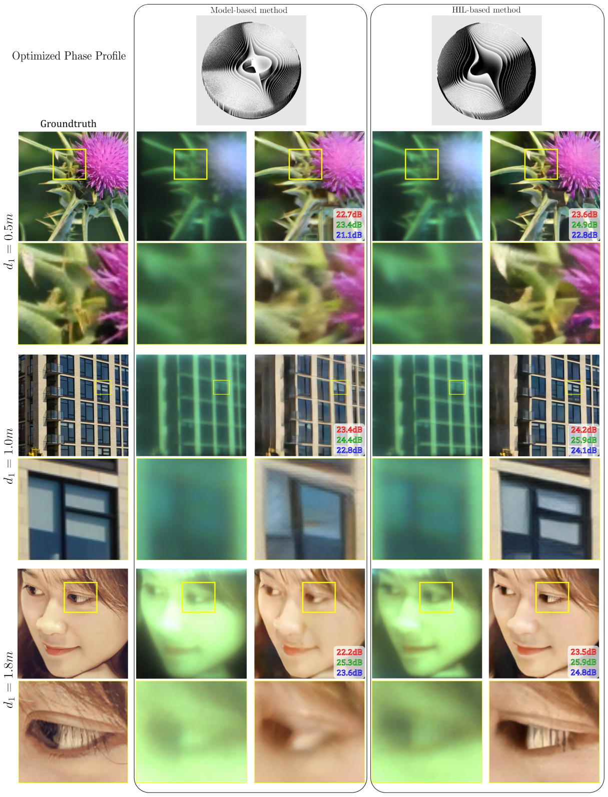

Figure 12. *Comparison of the performance of the designed DOE in 3 monitor setup hybrid imaging over testing dataset for two different end-to-end optimization approaches: model-based and HIL methodology. For each of the designed systems, a deep UNet-style CNN image reconstruction approach is employed (columns 3 and 5) to recover the blurred image on the sensor. The obtained images are presented in rows 1, 3, and 5 with their enlarged fragments in rows 2, 4, and 6 corresponding to the monitor distances from the camera $d_1 = 0.5, 1.0,$ and 1.8 m, respectively. The PSNR values over RGB channels are reported for each depth separately after image reconstruction. The uniform and high-quality imaging is achieved over depths and colors with PSNR values of about 25 dB for the hybrid setup optimized by the HIL methodology. As a counterpart, this value is about 24 dB for the hybrid setup with a phase pattern achieved by an analytical end-to-end optimization.*

Acknowledgments

This work is supported by the CIWIL project funded by Jane and Aatos Erkko Foundation, Finland.

References

[1] Lévêque, Olivier, et al. ”Co-designed annular binary phase masks for depth-of-field extension in single-molecule localization microscopy.” Optics Express 28.22 (2020): 32426-32446.

[2] Baek, Seung-Hwan, et al. ”End-to-end hyperspectral-depth imaging with learned diffractive optics.” (2020).

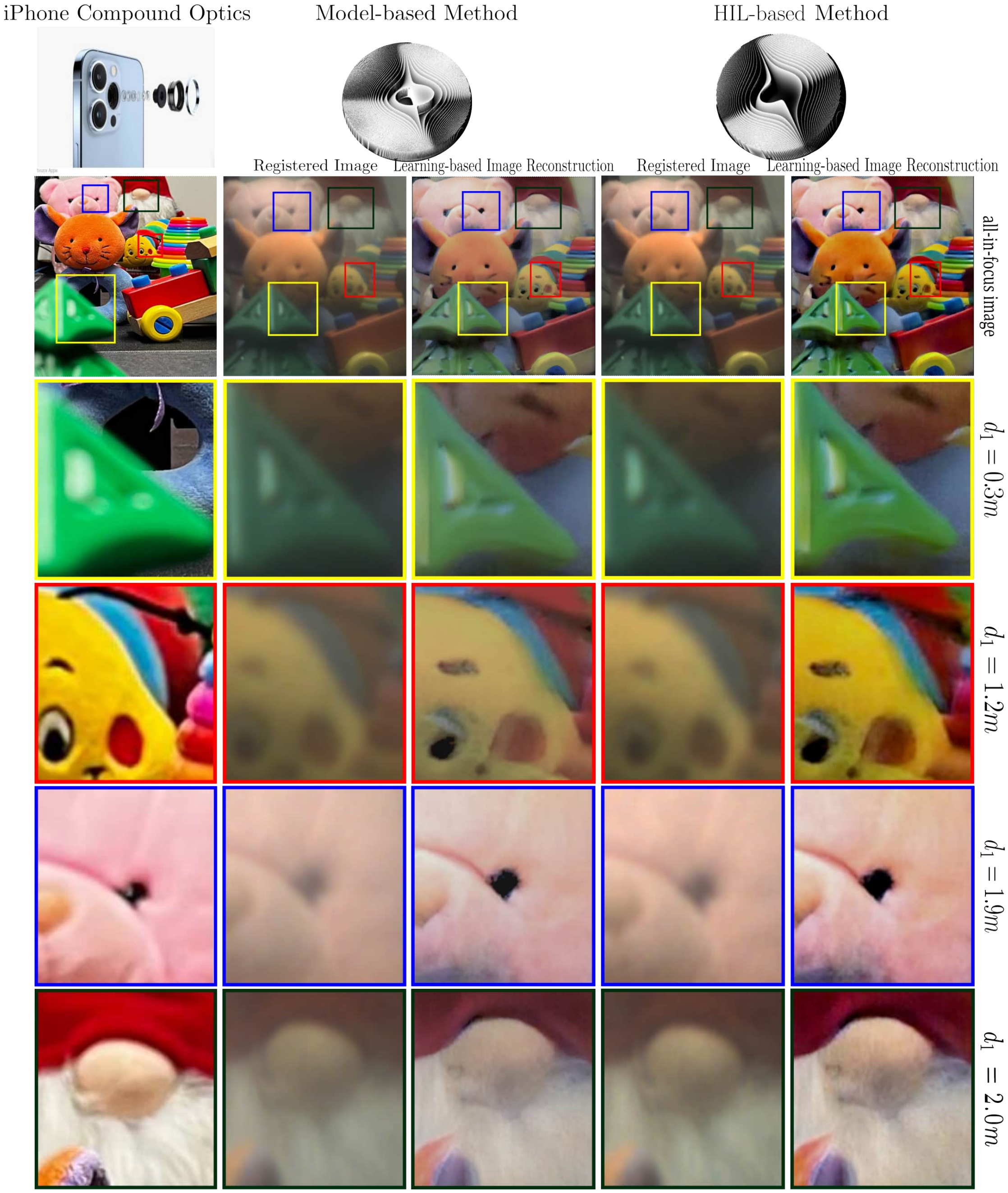

Figure 13. *Comparison of the designed DOE in real-scene scenario hybrid imaging for two different end-to-end optimization approaches: model-based and HIL methodology versus the compound lens camera of smartphone. For each of the designed systems, a deep UNet-style CNN image reconstruction approach is employed and trained separately (columns 3 and 5) to recover the blurred image on the sensor (columns 2 and 4). The obtained images are presented in row 2 with their enlarged fragments in rows 3, 4, 5, and 6 corresponding to four off-focus distances from the camera $d_1 = 0.3, 1.2, 1.9,$ and 2.0 m, respectively. By comparing the imaging results after reconstruction over the end-to-end optimization approaches for the hybrid design (columns 3 and 5), the advantage of the HIL-based method is clear. Overall, uniform and high-quality imaging is achieved over depths and colors. For the smartphone camera, the imaging quality is not good for close and far distances. The visual advantage in uniform sharpness and color preservation is clearly in favor of the designed hybrid imaging.*

[3] MiriRostami, SeyyedReza, Vladimir Y. Katkovnik, and Karen O. Eguiazarian. ”Extended DoF and achromatic inverse imaging for lens and lensless MPM camera based on Wiener filtering of defocused OTFs.” Optical Engineering 60.5 (2021): 051204.

[4] Chen, Wei Ting, Alexander Y. Zhu, and Federico Capasso. ”Flat optics with dispersion-engineered metasurfaces.” Nature Reviews Materials 5.8 (2020): 604-620.

[5] Tseng, E., Colburn, S., Whitehead, J., Huang, L., Baek, S. H., Majumdar, A., Heide, F. (2021). Neural nano-optics for high-quality thin lens imaging. Nature communications, 12(1), 1-7.

[6] Bayati, Elyas, et al. ”Inverse designed extended depth of focus meta-optics for broadband imaging in the visible.” Nanophotonics 11.11 (2022): 2531-2540.

[7] Colburn, Shane, Alan Zhan, and Arka Majumdar. ”Metasurface optics for full-color computational imaging.” Science advances 4.2 (2018): eaar2114.

[8] Whitehead, James EM, et al. ”Fast extended depth of focus meta-optics for varifocal functionality.” Photonics Research 10.3 (2022): 828-833.

[9] Sitzmann, Vincent, et al. ”End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging.” ACM Transactions on Graphics (TOG) 37.4 (2018): 1-13.

[10] Dun, Xiong, et al. ”Learned rotationally symmetric diffractive achromat for full-spectrum computational imaging.” Optica 7.8 (2020): 913-922.

[11] Krajancich, Brooke, Nitish Padmanaban, and Gordon Wetzstein. ”Factored occlusion: Single spatial light modulator occlusion-capable optical see-through augmented reality display.” IEEE transactions on visualization and computer graphics 26.5 (2020): 1871-1879.

[12] Jeon, Daniel S., Seung-Hwan Baek, Shinyoung Yi, Qiang Fu, Xiong Dun, Wolfgang Heidrich, and Min H. Kim. ”Compact snapshot hyperspectral imaging with diffracted rotation.” (2019).

[13] Adams, Jesse K., et al. ”Single-frame 3D fluorescence microscopy with ultraminiature lensless FlatScope.” Science advances 3.12 (2017): e1701548.

[14] Antipa, N., Kuo, G., Heckel, R., Mildenhall, B., Bostan, E., Ng, R., & Waller, L. (2018). DiffuserCam: lensless single-exposure 3D imaging. Optica, 5(1), 1-9.

[15] Yanny, Kyrollos, et al. ”Miniature 3D fluorescence microscope using random microlenses.” Optics and the Brain. Optica Publishing Group, 2019.

[16] Rostami, Seyyed Reza Miri, et al. ”Power-balanced hybrid optics boosted design for achromatic extended depth-of-field imaging via optimized mixed OTF.” Applied Optics 60.30 (2021): 9365-9378.

[17] Pinilla, S., Rostami, S. R. M., Shevkunov, I., Katkovnik, V., & Egiazarian, K. (2022). Hybrid diffractive optics design via hardware-in-the-loop methodology for achromatic extended-depth-of-field imaging. Optics Express, 30(18), 32633-32649.

[18] Seyyed Reza Miri Rostami, Samuel Pinilla, Igor Shevkunov, Vladimir Katkovnik, and Karen Eguiazarian, On design of hybrid diffractive optics for achromatic extended depth-of-field (EDoF) RGB imaging, Unconventional optical imaging III, 2022, pp. 160–175.

[19] Joseph W Goodman, Introduction to fourier optics, Roberts and Company Publishers, 2005.

[20] Katkovnik, Vladimir, Mykola Ponomarenko, and Karen Egiazarian. ”Lensless broadband diffractive imaging with improved depth of focus: wavefront modulation by multilevel phase masks.” Journal of Modern Optics 66.3 (2019): 335-352.

[21] Emmerich, Michael, and André H. Deutz. ”A tutorial on multiobjective optimization: fundamentals and evolutionary methods.” Natural computing 17.3 (2018): 585-609.

[22] Zhang, Kai, et al. ”Plug-and-play image restoration with deep denoiser prior.” IEEE Transactions on Pattern Analysis and Machine Intelligence (2021).

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

[24] Khan, Salman Siddique, et al. ”Flatnet: Towards photorealistic scene reconstruction from lensless measurements.” IEEE Transactions on Pattern Analysis and Machine Intelligence (2020).

[25] Goodfellow, Ian, et al. ”Generative adversarial networks.” Communications of the ACM 63.11 (2020): 139-144.

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

[27] Dabov, Kostadin, et al. ”Image denoising by sparse 3-D transform-domain collaborative filtering.” IEEE Transactions on image processing 16.8 (2007): 2080-2095.

[28] Hansen, Nikolaus, and Andreas Ostermeier. ”Adapting arbitrary normal mutation distributions in evolution strategies: The covariance matrix adaptation.” Proceedings of IEEE international conference on evolutionary computation. IEEE, 1996.

Author Biography

Seyyed Reza Miri Rostami *received his BSc degree in electronics engineering from Babol Noshirvani University of Technology, Babol, Iran, in 2015, and his MSc degree in electrical communication engineering from Tarbiat Modares University, Tehran, Iran, in 2017. He is currently a Ph.D. candidate at Tampere University, Finland. He has published ten journals and conference papers. His current research interests include image processing, optimization, parallel computing, signal processing, computational and diffractive optics, Hybrid, and lensless imaging.*

Samuel Pinilla *is currently research associate at University of Manchester. He received the B.S. degree (cum laude) in Computer Science in 2014, the B.S. degree in Mathematics, and the M.S degree in Mathematics from Universidad Industrial de Santander, Bucaramanga, Colombia in 2016 and 2017, respectively. His Ph.D. degree from the Department of the Electrical and Computer Engineering, Universidad Industrial de Santander, Bucaramanga, Colombia. His research interests focuses on the areas of high-dimensional structured signal processing and (non)convex optimization methods.*

Igor Shevkunov *is a postdoctoral researcher at Tampere University since 2017. He received his Ph.D. in Optics from St.Petersburg State University, Russia, in 2013. He is the author of more than 50 refereed papers. His main research interests are digital holography, phase retrieval, denoising, and interferometry.*

Vladimir Katkovnik *received his Ph.D. and DSc degrees in technical cybernetics from Leningrad Polytechnic Institute (LPI) in 1964 and 1974, respectively. From 1964 to 1991, he was an associate professor and then a professor in the Department of Mechanics and Control Processes, LPI.*

Since 2003, he has been with the Department of Signal Processing, Tampere University of Technology (TUT), Finland. He has published six books and over 350 refereed journal and conference papers. His research interests include stochastic image/signal processing, nonparametric estimation, computational imaging, and computational phase imaging.

***Karen Egiazarian** received his MSc degree from Yerevan State University in 1981, his Ph.D. from Moscow State University, Russia, in 1986, and his DTech degree from TUT, Finland, in 1994. He is a professor leading the Computational Imaging Group, ICT faculty, Tampere University. He has authored about 650 refereed journal and conference papers. His research interests include computational imaging, sparse coding, and image and video restoration. He serves as an associate editor for the IEEE Transactions of Image Processing and is the editor-in-chief of the Journal of Electronic Imaging.*