Research Article | Vol. 30, No. 22 / 24 Oct 2022 / Optics Express 40871
Optics EXPRESS

High-performance full-color imaging system based on end-to-end joint optimization of computer-generated holography and metalens

ZEQING YU,[1] QIANGBO ZHANG,[1] XIAO TAO,[2] YONG LI,[1,3] CHENNING TAO,[1] FEI WU,[3] CHANG WANG,[1,4,5] AND ZHENRONG ZHENG[1,4,*]

[1]State Key Laboratory of Modern Optical Instrumentation, College of Optical Science and Engineering, Zhejiang University, Hangzhou 310027, China
[2]Shanghai Aerospace Control Technology Institute, Shanghai 201109, China
[3]Beijing LLVision Technology Co., Ltd., Room 301, Building B12C, No. 10 Jiuxianqiao Rd, Chaoyang District, Beijing 10015, China
[4]Intelligent Optics & Photonics Research Center, Jiaxing Research Institute Zhejiang University, Jiaxing 314000, China
[5]changwang_optics@zju.edu.cn
*zzr@zju.edu.cn

Abstract: Metasurface has drawn extensive attention due to its capability of modulating light with a high degree of freedom through ultrathin and sub-wavelength optical elements, and metalens, as one of its important applications, promises to replace the bulky refractive optics, facilitating the imaging system light-weight and compact characteristics. Besides, computer-generated holography (CGH) is of substantial interest for three-dimensional (3D) imaging technology by virtue of its ability of restoring the whole optical wave field and re-constructing the true 3D scene. Consequently, the combination of metalens and CGH holds transformative potential in enabling the miniaturization of 3D imaging systems. However, its imaging performance is subject to the aberrations and speckle noises originating from the metalens and CGH. Inspired by recent progress that computational imaging can be applied to close the gap, a novel full-color imaging system, adopting end-to-end joint optimization of metalens and CGH for high imaging quality, is proposed in this paper. The U-net based network as the pre-processing adjusts weights to make the holographic reconstruction offset imaging defects, incorporating the imaging processing into the step of generating hologram. Optimized by deep learning, the proposed imaging system is capable of full-color imaging with high fidelity in a compact form factor, envisioned to take an essential step towards the high-performance miniaturized imaging system.

1. Introduction

The miniaturization of the imaging system has presented considerable potential in various applications such as virtual and augmented reality, medical imaging, commodity smartphones, wearable and portable devices, and so on. However, the task of further scaling down the dimensions and weight of the system while retaining its functionality is challenging, because the conventional optics have to be cascaded and bulky to meet the high imaging requirements. Nowadays, metasurface has emerged as ultrathin optics for accurately modifying the wavefront with large design freedom [1–4]. This remarkable ability enables the metasurface unprecedented functionalities, one of which, metalens, paves the way for further miniaturizing the imaging system as an alternative option to conventional diffractive optics [5–10].

In addition, the computer-generated holography (CGH) has been widely investigated as a key technology for three-dimensional (3D) imaging attributed to its ability of recovering the

#470419 https://doi.org/10.1364/OE.470419
Journal © 2022 Received 20 Jul 2022; revised 21 Sep 2022; accepted 5 Oct 2022; published 20 Oct 2022

Optics EXPRESS

true 3D scene [11–15]. Recently, the CGH competent in full-color 3D display [16,17], can collaborate with metalens to form a compact 3D holographic display system [8]. However, the phase discontinuities of the metalens and structural imperfections of holographic devices result in strong aberration and speckle noises [18–20], which is especially worse in full-color imaging, degrading imaging quality.

To tackle the aforementioned barriers, we turn towards the emerging computational imaging which opens up new directions in improving imaging quality [7,19,21–23], shifting the load from hardware to algorithms. It has played an important role in a wide range of fields, such as wide field-of-view imaging [24], depth prediction [25], snapshot hyperspectral imaging [26], aberration correction [19,21,24,27], and so forth. Lately, post-processing method based on end-to-end deep learning, which jointly optimizes the imaging optics and imaging processing approach, holds the latent capacity in refining the imaging performance of the imaging system, stimulating progress in high-quality imaging optics drastically [28–34]. However, post-processing approach is unable to obtain images directly, imposing restrictions on some specific applications such as the near-eye display.

Inspired by prior work, a metalens-based holographic full-color imaging system exploiting the end-to-end joint optimization of neural network and metalens for high-performance imaging is proposed in this paper, which is shown schematically in Fig. 1(a). In the proposed system, the RGB laser sources are incident on the holographic device, spatial light modulator (SLM). After being modulated in accordance with the hologram loaded on the SLM, the laser sources are reflected and diffracted to form a holographic reconstructed image (HRI), which is imaged by the metalens (circularly polarized polarizers are needed in practice, which is introduced in detail later and omitted here) and detected by the sensor later. The network is trained to generate the hologram, which is optimized simultaneously with the phase profile of the metalens. By tuning the weights of the network and the phase profile of the metalens, the HRI is produced to be deliberately spatially non-conforming to the ground-truth image in order to offset the aberration, making the final imaging result as close to the ground-truth image as possible. The synergistic combination of the metalens, CGH and computational imaging offers compact form factors to imaging system without compromising imaging quality, further boosting the potential applications of compact imaging devices in wearable and portable platforms.

Specifically, the contributions of the proposed system are introduced as follows.

1. The most adopted post-processing is replaced by the pre-processing method which avoids additional processing after sensor imaging and stimulates the possible applications in direct imaging.

2. Introducing the metalens and the CGH endows the imaging system with a compact form and 3D display potential, with the capability to overcome the challenges of focus cues, vergence-accommodation conflicts [35], vision correction [12], etc.

3. The system integrates the pre-processing with the hologram-generated network to alleviate the complexity of the system. The network not only generates the hologram but also recodes the phase information of the hologram to overcome the display defects from CGH and the metalens.

Optics EXPRESS

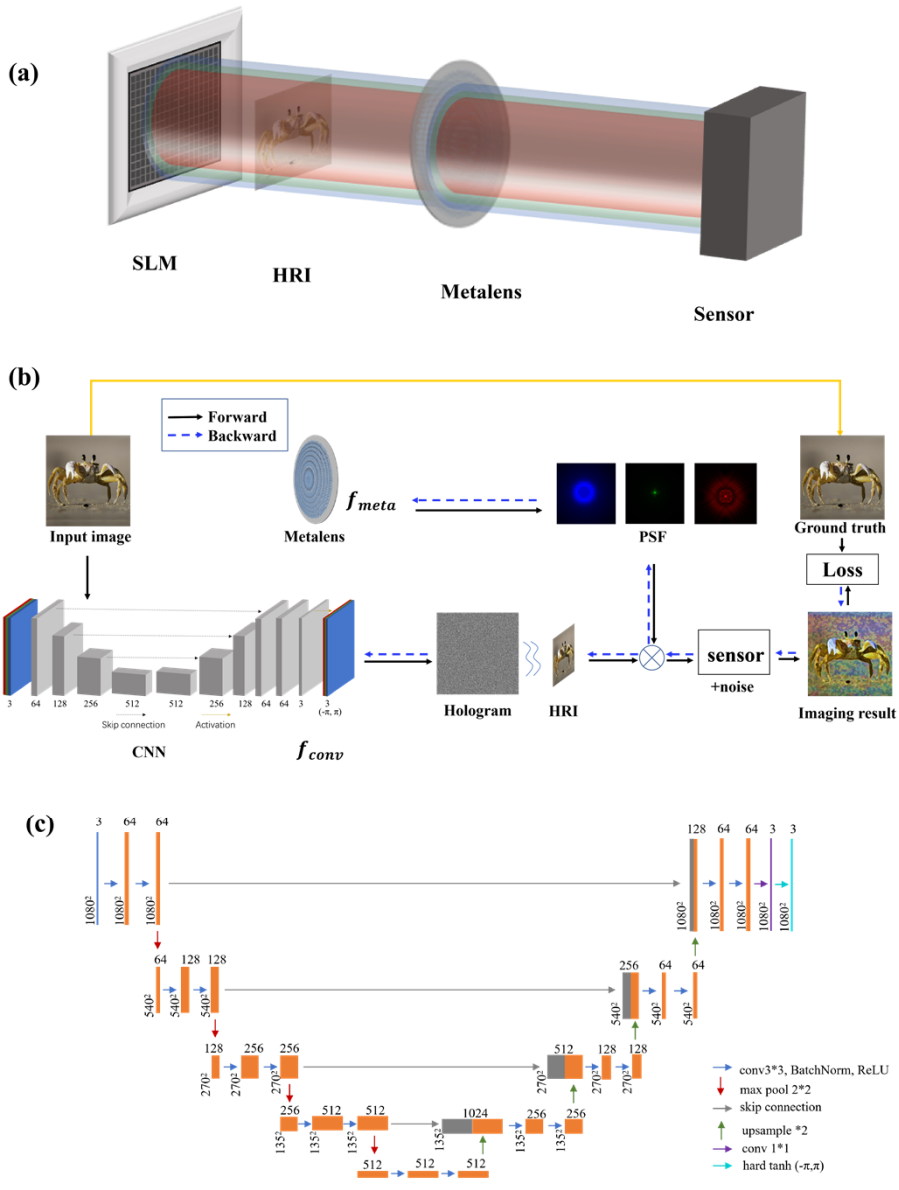

Fig. 1. (a) Schematic diagram of the proposed holographic meta-based imaging system. (b) End-to-end imaging pipeline based on joint optimization of the metalens and network. (c) The detailed multilayered structure of the hologram-generated neural network.

Optics EXPRESS

2. Theoretical analyses

The pipeline of the optimizing procedure exhibited in Fig. 1(b) involves five main stages: First, the neural network takes the RGB image as the input and outputs the holographic phase pattern, followed by diffractive propagation to form HRI. Second, the phase profile of metalens is the sum of an initial phase and a polynomial, which is decided by a set of optimizable coefficients. Third, deduce the point spread function (PSF) from the phase profile of metalens. Fourth, the imaging process is implemented through convolving the PSF with HRI and adding noise to the convolution result. Finally, the loss between the recovery and the ground truth is calculated for backward propagation in order to update the trainable variables.

2.1. Network structure

A U-net [36] based convolutional neural network (CNN), which performs a multi-scale operation on the image, is developed to learn the map relating the target amplitude distribution to the hologram. The architecture of the network is demonstrated in detail in Fig. 1(c). The input layer of 3 channels is connected with an initial layer of 64 channels, followed by four down-sampling and up-sampling blocks. After taking a max-pooling layer and a bilinear up-sampling layer and a bilinear up-sampling layer as the beginning, respectively, both the down-sampling and up-sampling blocks include two convolutional layers, two normalization layers and two activation layers with leaky ReLU. Through skip connections between the up-sampling and down-sampling blocks, the scaled features are concatenated with upper-scale features. The last convolutional layer outputs the result with 3 channels, following a hard tanh function mapping the value between [-π, π] to produce the phase pattern of RGB-color holograms.

2.2. Holographic principle

The operation of the phase-generating network is referred to as f_{CNN}, thus the phase pattern is derived as

$$\varphi_{\mathrm{h}}(x_{\mathrm{h}}, y_{\mathrm{h}}) = f_{\mathrm{CNN}}(\theta_{\mathrm{CNN}}, \sqrt{I_{\mathrm{truth}}}), \tag{1}$$

where the θ_{CNN} represents the whole trainable parameters of the network and I_{truth} means the intensity of the ground truth image. Since the simultaneous modulation of both amplitude and phase is prohibited by the limitation of SLM, merely the phase-only modulation is implemented for CGH here. Therefore, the complex amplitude of the hologram plane is written as

$$u_{\mathrm{h}}(x_{\mathrm{h}}, y_{\mathrm{h}}) = \exp\left(i\varphi_{\mathrm{h}}(x_{\mathrm{h}}, y_{\mathrm{h}})\right). \tag{2}$$

To form the HRI from the hologram, a diffractive propagation method called band-limited angular spectrum method (BL-ASM) [37] is exploited, and the principles of which are summarily introduced here. The notation

$$u_{\mathrm{rec}} = f_{\mathrm{BL\_ASM}}(u_{\mathrm{h}}) \tag{3}$$

indicates the transformation of BL-ASM, of which the intuitive interpretation is depicted in Fig. 2 and the detailed process is explained as follows. First, as Fig. 2(a) shows, the input $u_{\mathrm{h}}(x_{\mathrm{h}}, y_{\mathrm{h}})$ is padded zeros to twice the size as $u'_{\mathrm{h}}(x_{\mathrm{h}}, y_{\mathrm{h}})$ in order to convert the circle convolution to the linear convolution. Then the corresponding transfer function of the expanded field is calculated as $H(f_x, f_y)$, which is defined as

$$H(f_x, f_y) = \begin{cases} \exp\left(\mathrm{i}\frac{2\pi}{\lambda} z\sqrt{1-(\lambda f_x)^2-(\lambda f_y)^2}\right), & \text{if } \sqrt{f_x^2+f_y^2} < \frac{1}{\lambda} \\ 0, & \text{otherwise} \end{cases}. \tag{4}$$

Optics EXPRESS

To avoid aliasing errors, $H(f_x, f_y)$ is bounded within thresholds along f_x and f_y directions as $H_T(f_x, f_y)$, which is given as

$$H_T(f_x,f_y) = H(f_x,f_y)\text{rect}(\frac{f_x}{2T_{fx}})\text{rect}(\frac{f_y}{2T_{fy}}). \tag{5}$$

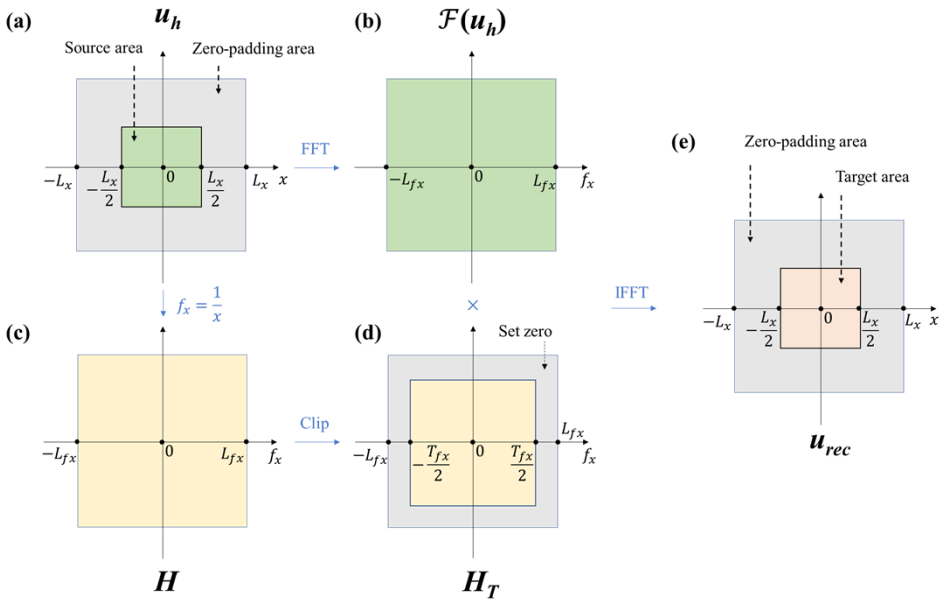

Fig. 2. Schematic of the BL-ASM. (a) Padding the input with zeros to the twice original sizes to convert the circle convolution to the linear convolution. (b) The Fourier transform of (a). (c) The corresponding frequency domain of the expanded field in (a). (d) The transfer function bounded to avoid aliasing errors. (e) The result clipped to the original size after being reconstructed by inverse Fourier transform of the product of fields in (b) and (d).

Here the thresholds are defined as

$$T_{fx}=\frac{1}{\lambda\sqrt{(2\Delta f_x z)^2+1}}, T_{fy}=\frac{1}{\lambda\sqrt{(2\Delta f_y z)^2+1}}. \tag{6}$$

The expressions of Δf_x and Δf_y mean the sampling spacing of the frequency domain coordinates f_x and f_y, respectively. Next, the whole process of BL-ASM can be expressed as

$$u'_{\text{rec}}(x_{\text{rec}}, y_{\text{rec}})=\mathcal{F}^{-1}\{\mathcal{F}[u'_{\text{h}}(x_{\text{h}}, y_{\text{h}})]H_{\text{T}}(f_x,f_y)\}, \tag{7}$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ mean Fourier transform and inverse Fourier transform, respectively. Finally, the result $u'_{\text{rec}}(x_{\text{rec}}, y_{\text{rec}})$ is cropped to the same size of the input as the holographic reconstructed image $u_{\text{rec}}(x_{\text{rec}}, y_{\text{rec}})$, as Fig. 2(e) shows.

2.3. Metalens design

To realize metalens design, Pancharatnam–Berry phase manipulation is selected from various nanophotonic design mechanisms for its high accuracy of phase compensation realization [18]. In this case, the nanofins serve as half-wave plates, transforming the helicity of incident circularly

Optics EXPRESS

polarized light with local phase shift by being rotated, which is written as

$$u_t = \frac{t_L + t_S}{2}|\sigma\rangle + \frac{t_L - t_S}{2}\exp(j2\sigma\beta)|-\sigma\rangle, \tag{8}$$

$$|\sigma\rangle = \frac{(1 + i\sigma)^T}{\sqrt{2}}. \tag{9}$$

The parameters t_L and t_S are the complex transmission coefficients for linear polarized beams along longer and shorter optical axes of the nanofin, and β is the orientation angle along z axis. Equation (9) represents the unit vector of circularly polarized light, where spin-charge $\sigma = 1$ and $\sigma = -1$ signify left and right circularly polarized light, respectively. When a left circularly polarized light is incident on the metalens, the transmitted light is combined of two components. The first and second terms in Eq. (8) denote the transmitted beam with the same and opposite circular polarization state of the incident light. Besides, the second part is imparted with an additional 2β phase shift, which can be filtered out by a circularly polarized polarizer. Supposing the nanofin is rotated from 0° to 180°, the phase manipulation can cover the whole 0-2π range. The proportion of the change of the polarization state is decided by the polarization conversion efficiency (PCE), which is defined as the power of transmitted right circularly polarized light I_{rcp} divided by the total incident power $I_{input:}$

$$\mathrm{PCE} = \frac{I_{rcp}}{I_{input}}. \tag{10}$$

The design of the metalens unit cell is depicted in Fig. 3(b). The unit cell is composed of quartz substrate with the lattice period $P = 400$ nm and Si_3N_4 nanofin with the width $W = 115$ nm, the length $L = 315$ nm and the height $H = 750$ nm. The nanostructure is optimized by the full wave finite-difference time-domain (FDTD) simulations, and the wavelength-dependent PCE can be obtained, which is plotted from 400 nm to 700 nm with the wavelength spacing of 10 nm in Fig. 3(d).

So, the metalens can be designed according to the phase distribution in need. Given that the completely random optimization of its phase profile is far away from the optimal solution and too computationally time-consuming, the phase profile of metalens is defined by both an initial phase and a polynomial as

$$\varphi_{meta}(x, y) = \varphi_{initial}(x, y) + \sum_{i=0}^{N} \theta_i \left(\frac{x^2 + y^2}{R^2}\right)^i, \tag{11}$$

where (x, y) is the spatial coordinate along the metalens, $\{\theta_0, \ldots, \theta_N\}$ is the collection of trainable variables, R is the radius of the metalens, and N is set to 8 as the number of terms. The initial phase used here is a hyperboloidal form [38] given as

$$\varphi_{initial}(x, y) = \frac{2\pi}{\lambda}(f - \sqrt{x^2 + y^2 + f^2}). \tag{12}$$

The planar incident light could be modulated by the metalens with the initial phase profile into a spherical wavefront, converged at the distance f, i.e., the focal length of the metalens. λ is set to be 520 nm as the nominal wavelength. At the beginning of the training, only the weights of the network are updated. The metalens phase is fixed at $\varphi_{initial}$ to make the network learn to generate the hologram. After a few training steps, the polynomial decided by trainable parameters $\{\theta_0, \ldots, \theta_N\}$ is appended to the definition of the metalens and evolved together with the network. At last, the network is finetuned according to the moderate changes of the phase profile of metalens.

Optics EXPRESS

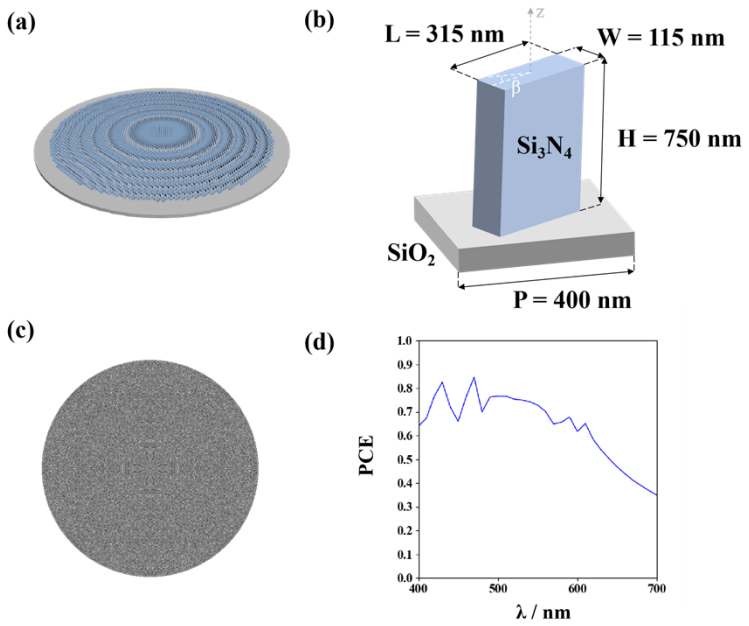

Fig. 3. (a) The outlook of the metalens. (b) The detailed structure of the metalens unit cell: Si_3N_4 nanofin on the SiO_2 substrate. (c) The optimized phase of the metalens. (d) PCE of the unit cell simulated in the FDTD from 400 nm to 700 nm with the wavelength spacing of 10 nm.

2.4. PSF simulation and imaging process

Since the holographic display relies on laser sources, the imaging system is modeled as a convolution of the image and PSF with complex amplitude under coherent illumination. Providing a coherent point light source at infinity illuminates the metalens, the PSF is the response of the complex wave field after transmitting metalens. The PSF is assumed nearly shift-invariant within a limited field of view and spatially symmetric for avoiding excessive computational cost, so it is determined mainly by the phase of the metalens and the incident wavelength, which is formulated as

$$\mathrm{PSF} = f_{\mathrm{PSF}}(\theta_{\mathrm{meta}}, \lambda). \tag{13}$$

The expression f_{PSF} generates the PSF from the metasurface phase parameters by Fresnel diffraction on the basis of the scalar diffraction theory [39]:

$$\mathrm{PSF} = \frac{\exp(i\lambda f)}{i\lambda f}\mathcal{F}\{u_{\mathrm{meta}}(x_{\mathrm{meta}}, y_{\mathrm{meta}})\exp(\frac{i\pi}{\lambda f}(x_{\mathrm{in}}^2 + y_{\mathrm{in}}^2))\}, \tag{14}$$

where $u_{\mathrm{meta}}(x_{\mathrm{meta}}, y_{\mathrm{meta}})$ is the complex wave field passing through the metalens, which is given as $u_{\mathrm{in}}(x_{\mathrm{in}}, y_{\mathrm{in}}) = A\exp(i\varphi_0)$ represents the complex amplitudes of the normalized RGB point sources at infinity with the amplitude A of 1 and the phase φ_0 of 0, and τ is the square root of PCE relative to the incident wavelength.

$$u_{\mathrm{meta}}(x_{\mathrm{meta}}, y_{\mathrm{meta}}) = \tau \cdot u_{\mathrm{in}}(x_{\mathrm{in}}, y_{\mathrm{in}})\exp(i\varphi_{\mathrm{meta}}) = \tau A\exp(i(\varphi_0 + \varphi_{\mathrm{meta}})) \tag{15}$$

The imaging process can be achieved by convolving the PSF with the complex wave field of HRI, which is written as

$$u_{\mathrm{imaging}} = \mathrm{PSF} * u_{\mathrm{rec}}. \tag{16}$$

Optics EXPRESS

In consideration of the noise influence, the sensing process is modeled as adding the noise function to the imaging result, which is the squared value of u_{imaging} :

$$I_{\text{sensor}} = |u_{\text{imaging}}|^2 + \eta = \left| f_{\text{meta}}(\theta_{\text{meta}}) * f_{\text{BL\_ASM}}(\exp(if_{\text{CNN}}(\theta_{\text{CNN}}, \sqrt{I_{\text{truth}}}))) \right|^2 + \eta. \tag{17}$$

η represents Gaussian noise added to enhance the sensitivity to noise in reconstruction.

2.5. Loss definition

During the training, the loss function is defined as a combination of ℓ_2 (mean squared loss) and perceptual losses [40,41] to evaluate the deviation of the recovered image with the ground truth, denoted as

$$loss = loss_{\text{mse}} + \lambda_{\text{p}} loss_{\text{p}}, \tag{18}$$

$$loss(I_{\text{sensor}}, I_{\text{truth}}) = ||I_{\text{sensor}} - I_{\text{truth}}||_2 + \lambda_{\text{p}} \sum_{l=1}^{L} ||P_l(I_{\text{sensor}}) - P_l(I_{\text{truth}})||_2. \tag{19}$$

The weight coefficient of perceptual loss λ_{p} is set as 0.01 here. And the function P_l is referred to as a VGG based perceptual loss function transforming the image to perceptual feature space [42]. The variable l enumerates the layers of the VGG network with the maximum value $L = 5$. As a routine, the training is carried out by back-propagating the loss function to adjust the optimizable parameters of the phase profile of the metalens and the network to minimize the training loss function, which is given as:

$$\arg\min_{\theta_{meta}, \theta_{CNN}} \sum_{j=1}^{J} loss(I^j_{sensor}, I^j_{truth}), \tag{20}$$

where J is the number of training samples.

3. Results and discussion

Particular simulations are conducted out to assess the performance of the proposed system. The diameter of the metalens is set as 1 mm with the sampling pitch of 400 nm, which is consistent with the period of the unit cell. And the focal length of the metalens is set as 15 mm. As for the light sources, the wavelengths of RGB laser beams are 450 nm, 520 nm and 660 nm, with the PCE values of 0.66, 0.75 and 0.44 according to the simulation results shown in Fig. 3(d), respectively.

The training dataset is 1080p images from the DIV2K dataset [43], transformed with randomly vertical and horizontal flipping and randomly cropped into the resolution of 1080*1080, matching the dimensions of the holograms. The trainable parameters of network and the phase profile of metalens are optimized using Adam optimizer [44] with the learning rates of 10^{-4} and 10^{-3} on GPU, respectively. At the beginning of the training, the weights of the network experience 40 evolutions within a training step. After 200 steps, the polynomial coefficients of the phase profile are updated 8 times, followed by the network updating 40 times in a training step.

The simulated PSFs after normalization are shown in Fig. 4(a). Figure 4(b) shows the cross-sectional intensity profiles of Fig. 4(a) along the x axis. It can be observed that the green light is focused perfectly while defocusing appears for the blue and red ones, which will degrade the imaging quality. Yet with the help of the network, the defocusing effect can be partially offset.

The parameters of the hologram are in agreement with the SLM, which refers to a specific product as HOLOEYE LETO−3 Phase Only Spatial Light Modulator with the resolution and the sampling interval of 1080*1920 and 6.4 μm. Thus, the resolution of the hologram is set 1080*1080 with the sampling pitch of 6.4 μm. The diffraction distance from the SLM is set 100 mm in the algorithm. For the sensor Gaussian noise $\eta \sim N(0, \sigma^2)$, the standard deviation σ is set as 10^{-3}.

Optics EXPRESS

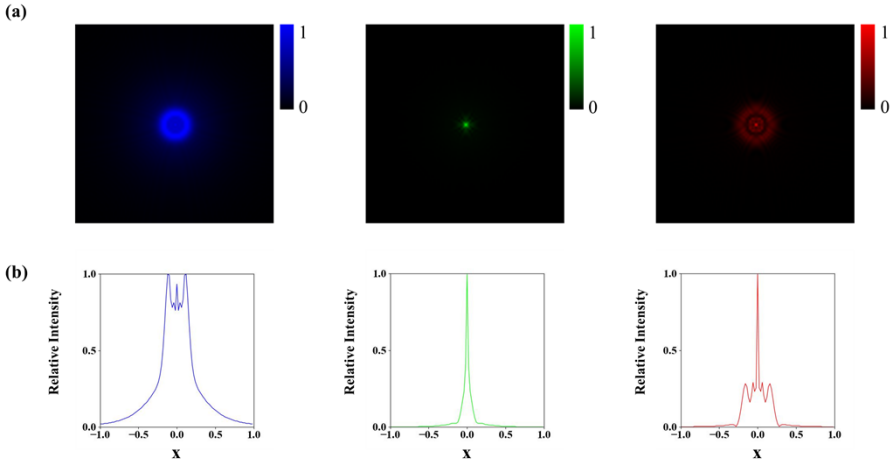

Fig. 4. (a) The relative intensity of the PSFs of the optimized metalens. (b) The cross-sectional intensity profiles along the *x*-axis at 0 of the *y*-axis.

The imaging performances of the imaging system are characterized by the imaging results exhibited in Fig. 5. The ground-truth images are in the left column. For comparisons, the images in the second column are generated using a metalens with the hyperboloid phase to observe the HRI based on the classical Gerchberg–Saxton (GS) iterative method [45] with BL-ASM. The results in the third column use the network to produce the holograms and the HRIs are observed by the metalens in the definition of the hyperboloid phase. And the last column exhibits the products of the proposed joint optimizing method. The local enlarged views of the all images are displayed on the right side. It is obvious that, as a result of the heavy intrinsic aberration of the imaging system, the images in the second column are too vague to distinguish. Though the imaging ability performs better in the third column compared to the second column, the proposed method adopted in the last column corrects the aberrations and maintains the image details, achieving the clearest imaging. Besides, the evaluation functions, peak signal-to-noise ratio (PSNR) and structural similarity (SSIM), are chosen for quantitative analysis of the image quality, which is signed at the bottom of the recovered images in Fig. 5, further confirming the proposed approach improves the imaging quality quantitatively.

Moreover, in order to verify the generalization of the proposed optimization model, a test group composed of 75 colorful images, which are not contained in the train dataset, is utilized to evaluate the three approaches mentioned above. The average values of PSNR and SSIM calculated from three sets of reconstructed images are displayed in Table 1. The data proves the joint optimization model is competent at enhancing system imaging performance.

Table 1. Average values of PSNR and SSIM for 75 test images

	GS + Hyperboloid	Net + Hyperboloid	Joint Optimization
PSNR	13.17	17.32	20.81
SSIM	0.2378	0.5984	0.8423

Overall, though the slight foggy appearance remains in the imaging results, due to the limited correction ability of overcoming the inherent imaging defects of the network, the system with joint optimization outperforms the comparison groups, from both visual perception and quantitative evaluation. Therefore, based on the fact that the simulated results successfully preserve fine

Optics EXPRESS

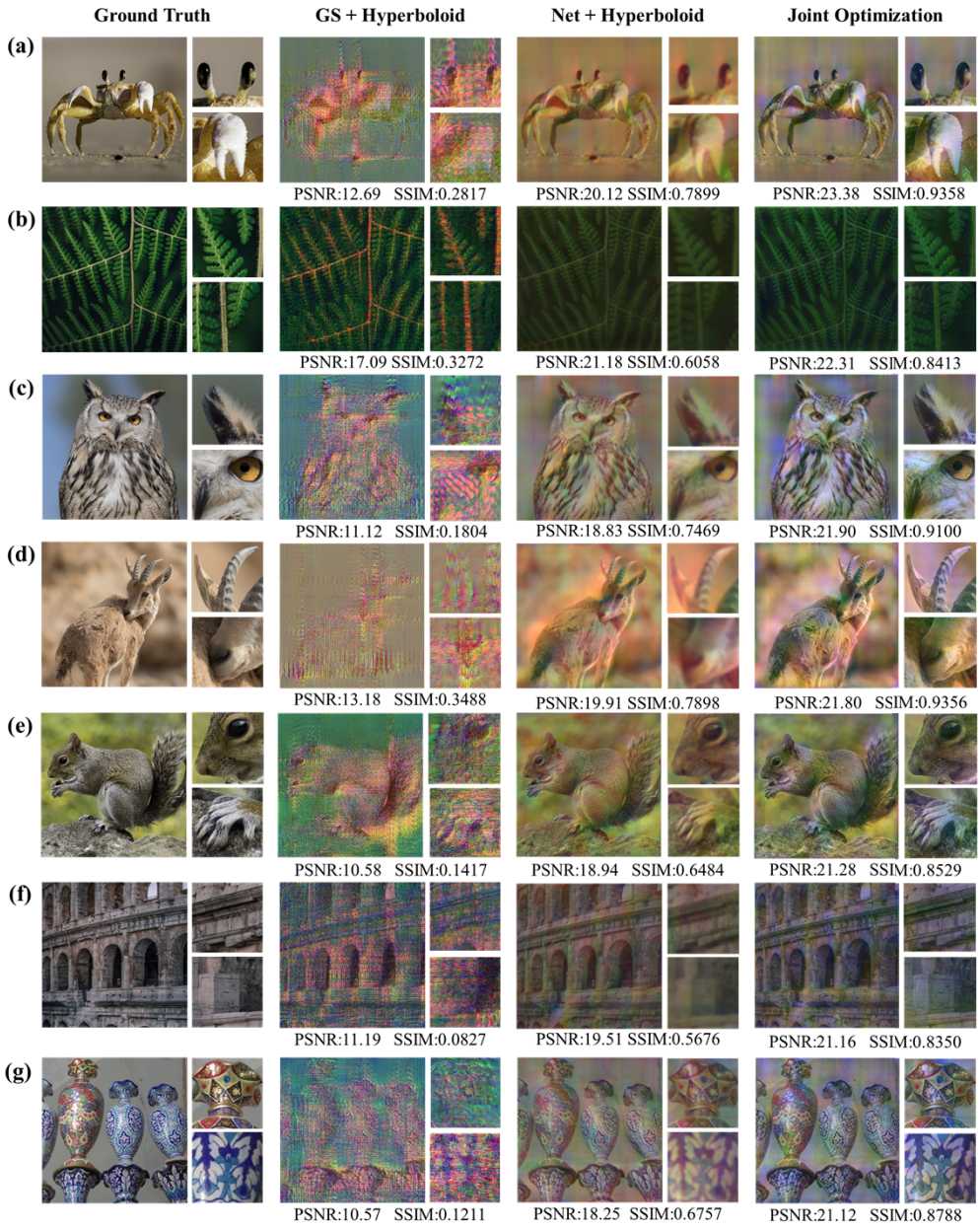

Fig. 5. The imaging results of proposed joint optimizing methods and comparison. The images in the first column are ground-truth images. The images in the second column for comparison are generated through the traditional GS holographic algorithm and imaging via metalens with the hyperboloid phase, which dramatically suffer from severe off-axis aberrations and chromatic aberration. The third and the last columns display the results using the network to generate holograms, while imaging through metalens with the hyperboloid phase and optimized phase, respectively. The enlarged local details and the quantitative evaluation results are displayed at the right and bottom of every recovery.

Optics EXPRESS

details and color fidelity, the effectiveness of eliminating aberrations using the proposed imaging system with a more compact form factor than traditional optical elements is validated.

Optics EXPRESS

4. Conclusion

In this paper, a novel full-color computational imaging system jointly optimizing the metalens and CGH for advanced imaging capabilities is proposed. In this case, the metalens with the feature of compactness, lightweight and subwavelength manipulation precision contributes to the miniaturization of the imaging system. Besides, the CGH technology imparts the system with 3D imaging potential to overcome the challenges of focus cues, vergence-accommodation conflicts, and so on. Moreover, leveraging the thriving computational optics theory, the system is enabled to optimize the hologram-generated network and the phase profile of metalens synergistically. The aberration correction is encoded through tuning the weights of the network. Thus, the reconstructed images of the holograms yielded from the network hold the potential of counteracting the imaging defects of the system. The simulated results verify the high imaging performance of the system, especially compared to the results without joint optimization. The proposed system, enabling imaging in a compact form without compromising quality, may facilitate the development of advanced imaging and display technologies in a wide range of applications requiring the compact factor.

Funding. Beijing Municipal Science and Technology Commission (Z201100004020012); Scientific Research Fund of Zhejiang Provincial Education Department (Y202148328).

Disclosures. The authors declare no conflicts of interest.

Data Availability. Data underlying the results presented in this paper are not publicly available but may be obtained from the authors upon reasonable request.

References

1. M. Khorasaninejad, W. T. Chen, R. C. Devlin, J. Oh, A. Y. Zhu, and F. Capasso, "Metalenses at visible wavelengths: Diffraction-limited focusing and subwavelength resolution imaging," Science **352**(6290), 1190–1194 (2016).
2. H. Zuo, D. Cho, X. Gai, P. Ma, and L. Xu, "High-Efficiency All-Dielectric Metalenses for Mid-Infrared Imaging," Adv. Opt. Mater. **5**(23), 1700585 (2017).
3. W. T. Chen, A. Y. Zhu, V. Sanjeev, M. Khorasaninejad, Z. Shi, E. Lee, and F. Capasso, "A broadband achromatic metalens for focusing and imaging in the visible," Nat. Commun. **13**(3), 220–226 (2018).
4. E. Arbabi, A. Arbabi, S. M. Kamali, Y. Horie, M. Faraji-Dana, and A. Faraon, "MEMS-tunable dielectric metasurface lens," Nat Commun **9**(1), 812 (2018).
5. J. Engelberg and U. Levy, "The advantages of metalenses over diffractive lenses," Nat. Commun. **11**(1), 1991 (2020).
6. D. Lin, P. Fan, E. Hasman, and M. L. Brongersma, "Dielectric gradient metasurface optical elements," Science **345**(6194), 298–302 (2014).
7. Y. Peng, Q. Sun, D. Xiong, G. Wetzstein, and F. Heide, "Learned large field-of-view imaging with thin-plate optics," ACM Trans. Graph. **38**(6), 1–14 (2019).
8. C. Wang, Z. Yu, Q. Zhang, Y. Sun, C. Tao, F. Wu, and Z. Zheng, "Metalens Eyepiece for 3D Holographic Near-Eye Display," Nanomaterials **11**(8), 1920 (2021).
9. G. Lee, J. Hong, S. Hwang, S. Moon, H. Kang, S. Jeon, H. Kim, J. Jeong, and B. Lee, "Metasurface eyepiece for augmented reality," Nat. Commun. **9**(1), 4562 (2018).
10. Z. Li, P. Lin, Y. Huang, J. Park, W. T. Chen, Z. Shi, C. Qiu, J. Cheng, and F. Capasso, "Meta-optics achieves RGB-achromatic focusing for virtual reality," Sci. Adv. **7**(5), eabe4458 (2021).
11. Q. Jiang, G. Jin, and L. Cao, "When metasurface meets hologram: principle and advances," Adv. Opt. Photonics **11**(3), 518–576 (2019).
12. Z. He, X. Sui, G. Jin, and L. Cao, "Progress in virtual reality and augmented reality based on holographic display," Appl. Opt. **58**(5), A74–A81 (2019).
13. S. A. Benton and V. Michael Bove Jr, *Holographic Imaging* (John Wiley & Sons, 2008).
14. D. Pi, J. Liu, and Y. Wang, "Review of computer-generated hologram algorithms for color dynamic holographic three- dimensional display," Light: Sci. Appl. **11**(1), 231 (2022).
15. D. Pi, J. Wang, J. Liu, J. Li, Y. Sun, Y. Yang, W. Zhao, and Y. Wang, "Color dynamic holographic display based on complex amplitude modulation with bandwidth constraint strategy," Opt. Lett. **47**(17), 4379–4382 (2022).
16. S. Choi, M. Gopakumar, Y. Peng, J. Kim, and G. Wetzstein, "Neural 3D holography," ACM Trans. Graph. **40**(6), 1–12 (2021).
17. L. Shi, B. Li, C. Kim, P. Kellnhofer, and W. Matusik, "Towards real-time photorealistic 3D holography with deep neural networks," Nature **591**(7849), 234–239 (2021).
18. N. Yu and F. Capasso, "Flat optics with designer metasurfaces," Nat. Mater. **13**(2), 139–150 (2014).
19. S. Colburn, A. Zhan, and A. Majumdar, "Metasurface Optics for Full-color Computational Imaging," Sci. Adv. **4**(2), eaar2114 (2018).

Optics EXPRESS

20. O. Avayu, E. Almeida, Y. Prior, and T. Ellenbogen, "Composite functional metasurfaces for multispectral achromatic optics," Nat. Commun. **8**(1), 14992 (2017).
21. F. Heide, M. Rouf, M. B. Hullin, B. Labitzke, and A. Kolb, "High-Quality Computational Imaging Through Simple Lenses," ACM Trans. Graph. **32**(5), 1–14 (2013).
22. K. Monakhova, J. Yurtsever, G. Kuo, N. Antipa, and L. Waller, "Learned reconstructions for practical mask-based lensless imaging," Opt. Express **27**(20), 28075–28090 (2019).
23. A. Sinha, J. Lee, S. Li, and G. Barbastathis, "Lensless computational imaging through deep learning," Optica **4**(9), 1117–1125 (2017).
24. E. Tseng, S. Colburn, J. Whitehead, L. Huang, S. Baek, A. Majumdar, and F. Heide, "Neural nano-optics for high-quality thin lens imaging," Nat. Commun. **12**(1), 6493 (2021).
25. H. Ikoma, C. M. Nguyen, C. A. Metzler, Y. Peng, and G. Wetzstein, "Depth from defocus with learned optics for imaging and occlusion-aware depth estimation," in *Proceedings of IEEE International Conference on Computational Photography*, (IEEE, 2021), pp. 1–12.
26. D. S. Jeon, S. H. Baek, S. Yi, Q. Fu, and H. K. Min, "Compact snapshot hyperspectral imaging with diffracted rotation," ACM Trans. Graph **38**(4), 1–13 (2019).
27. S. Colburn and A. Majumdar, "Simultaneous achromatic and varifocal imaging with quartic metasurfaces in the visible," ACS Photonics **7**(1), 120–127 (2020).
28. G. Barbastathis, A. Ozcan, and G. Situ, "On the use of deep learning for computational imaging," Optica **6**(8), 921–943 (2019).
29. V. Sitzmann, S. Diamond, Y. Peng, X. Dun, S. Boyd, W. Heidrich, F. Heide, and G. Wetzstein, "End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging," ACM Trans. Graph. **37**(4), 1–13 (2018).
30. J. Chang and G. Wetzstein, "Deep Optics for Monocular Depth Estimation and 3D Object Detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2019), pp. 10193–10202.
31. Y. Wu, V. Boominathan, H. Chen, A. Sankaranarayanan, and A. Veeraraghavan, "Phasecam3d — learning phase masks for passive single view depth estimation," in *Proceedings of IEEE International Conference on Computational Photography*, (IEEE, 2019), pp. 1–12.
32. C. Metzler, H. Ikoma, Y. Peng, and G. Wetzstein, "Deep optics for single-shot high-dynamic-range imaging," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*," (IEEE, 2020), pp. 1375–1385.
33. J. Chang, V. Sitzmann, X. Dun, W. Heidrich, and G. Wetzstein, "Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification," Sci. Rep. **8**(1), 12324 (2018).
34. X. Dun, H. Ikoma, G. Wetzstein, Z. Wang, X. Cheng, and Y. Peng, "Learned rotationally symmetric diffractive achromat for full-spectrum computational imaging," Optica **7**(8), 913–922 (2020).
35. D. M. Hoffman, A. R. Girshick, K. Akeley, and M. S. Banks, "Vergence-accommodation conflicts hinder visual performance and cause visual fatigue," Journal of Vision **8**(3), 33 (2008).
36. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer, 2015) pp. 234–241.
37. K. Matsushima and T. Shimobaba, "Band-Limited Angular Spectrum Method for Numerical Simulation of Free-Space Propagation in Far and Near Fields," Opt. Express **17**(22), 19662–19673 (2009).
38. F. Aieta, P. Genevet, M. A. Kats, N. Yu, R. Blanchard, Z. Gaburro, and F. Capasso, "Aberration-Free Ultrathin Flat Lenses and Axicons at Telecom Wavelengths Based on Plasmonic Metasurfaces," Nano Lett. **12**(9), 4932–4936 (2012).
39. J. W. Goodman, *Introduction to Fourier Optics,* 3*rd*, (Roberts and Company Publishers, 2004).
40. J. Johnson, A. Alahi, and F. F. Li, "Perceptual Losses for Real-Time Style Transfer and Super-Resolution," in *European conference on computer vision* (Springer, 2016), pp. 694–711.
41. Y. Peng, S. Choi, N. Padmanaban, J. Kim, and G. Wetzstein, "Neural Holography," in *ACM SIGGRAPH Emerging Technologies* (2020), pp. 1–2.
42. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556 (2014).
43. E. Agustsson and R. Timofte, "Ntire 2017 challenge on single image super-resolution: Dataset and study," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (IEEE, 2017), pp. 126–135.
44. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," https://arxiv.org/abs/1412.6980.
45. R. W. Gerchberg, "A practical algorithm for the determination of phase from image and diffraction pictures," Optik **35**(2), 1 (1972).