Research Article

ADVANCED PHOTONICS

Ultra-wide FOV meta-camera with transformer-neural-network color imaging methodology

Yan Liu,[a,†] Wen-Dong Li,[b,†] Kun-Yuan Xin,[a,†] Ze-Ming Chen,[a] Zun-Yi Chen,[a] Rui Chen,[a] Xiao-Dong Chen,[a] Fu-Li Zhao,[a] Wei-Shi Zheng,[b,*] and Jian-Wen Dong[a,*]

[a]Sun Yat-sen University, School of Physics, State Key Laboratory of Optoelectronic Materials and Technologies, Guangzhou, China
[b]Sun Yat-sen University, School of Computer Science and Engineering, Guangzhou, China

Abstract. Planar cameras with high performance and wide field of view (FOV) are critical in various fields, requiring highly compact and integrated technology. Existing wide FOV metalenses show great potential for ultrathin optical components, but there is a set of tricky challenges, such as chromatic aberrations correction, central bright speckle removal, and image quality improvement of wide FOV. We design a neural meta-camera by introducing a knowledge-fused data-driven paradigm equipped with transformer-based network. Such a paradigm enables the network to sequentially assimilate the physical prior and experimental data of the metalens, and thus can effectively mitigate the aforementioned challenges. An ultra-wide FOV meta-camera, integrating an off-axis monochromatic aberration-corrected metalens with a neural CMOS image sensor without any relay lenses, is employed to demonstrate the availability. High-quality reconstructed results of color images and real scene images at different distances validate that the proposed meta-camera can achieve an ultra-wide FOV (>100 deg) and full-color images with the correction of chromatic aberration, distortion, and central bright speckle, and the contrast increase up to 13.5 times. Notably, coupled with its compact size (< 0.13 cm^3), portability, and full-color imaging capacity, the neural meta-camera emerges as a compelling alternative for applications, such as micro-navigation, micro-endoscopes, and various on-chip devices.

Keywords: metalens; meta-camera; wide field of view; neural network; color imaging; compact imaging device.

Received Apr. 9, 2024; accepted for publication Apr. 18, 2024; published online May 20, 2024.

[DOI: 10.1117/1.AP.6.5.056001]

1 Introduction

Conventional cameras are renowned for their large imaging field of view (FOV) and unparalleled image quality. Due to the use of complex optical components for aberrations correction, it has a bulky architecture and faces the challenges of high-precision alignment. With the advancement of technology, the miniaturization, light weight, and portable cameras[1–3] are increasingly desired in autonomous driving, endoscopic medicine, and consumer electronics. Therefore, there is an urgent need for planar and high-performance optical components to implement wide FOV on-chip cameras.

Recently, metalenses composed of subwavelength artificial structures have garnered attention for their compactness, as potential alternatives to bulky and complex optical instruments.[4–9] The metalens exhibits superior optical performance due to its ability to precisely manipulate the incidence beam;[10–15] however, aberrations correction still remains a challenge, in particular for chromatic aberration and off-axis monochromatic aberration. To eliminate chromatic aberration, the dispersive propagation phase and dispersive-free geometric phase have been introduced to achieve broadband[16–19] and multiwavelength[20–22] achromatic metalenses. Due to the limitation of the group delay dispersion of meta-atoms, achromatic metalenses are usually implemented on the paraxial condition. In the off-axis case, a great deal of effort has been made to correct the off-axis monochromatic aberrations of the metalens. By introducing the ray-tracing method,[23–27] Fourier analysis,[28] and metalens array,[29] cascaded metalenses located on either side of the substrate[23–25] and single

*Address all correspondence to Wei-Shi Zheng, zhwshi@mail.sysu.edu.cn; Jian-Wen Dong, dongjwen@mail.sysu.edu.cn

†These authors contributed equally to this work.

wide FOV metalens with an aperture[26,27] can correct off-axis monochromatic aberrations and achieve diffraction-limited imaging over a wide FOV. However, chromatic aberration correction, central bright speckle, and image quality improvement of wide FOV metalens in the off-axis case are rarely considered. Obviously, addressing the aforementioned issues of wide FOV metalenses that rely on the above existing methods remains a considerable challenge.

To improve the image quality of a metalens, traditional image restoration computational imaging methods[2,22,30–32] are introduced, and they usually recover images based on simple hypotheses or enhance images through multiple-image super-resolution.[2,22,30] However, the factors influencing the imaging quality of current ultra-wide FOV metalenses are intricate, making it difficult to improve the imaging quality based on a single hypothesis. In recent years, several methods have been proposed to incorporate neural networks for improving the imaging quality of metalenses or diffractive optical elements, which use point spread functions (PSFs) to train models.[33–37] Unfortunately, single-wavelength ultra-wide FOV metalenses have complex PSF spatial variations in different incident angles at other non-designed wavelengths, which makes it fail to accurately model the imaging degradation by applying the above PSF method. Even worse, the central bright speckle indicates that there is inconsistency between the simulation data and the actual scene, which makes it more difficult to improve the imaging quality of ultra-wide FOV metalenses by the prior PSF method.

With the advancement of deep-learning[38] research, transformer modules based on attention mechanisms have been developed and demonstrated to be effective in cutting-edge studies, such as AlphaFold2,[39] GPT,[40] and large image-text models. Compared to CNN networks constructed with local convolutional kernels, the multi-head self-attention mechanism enables the transformer module to effectively model long-range dependencies, which is conducive to better modeling of wide FOV metalenses' non-focused diffusion spots and information expansion problems. It is expected that incorporating transformer methodology into wide FOV metalens imaging is a good choice to cope with more complex PSF spatial variations so as to largely improve the quality of imaging.

In this work, we demonstrate a highly miniaturized neural meta-camera in conjunction with an ultra-wide FOV metalens assembled on a CMOS image sensor. The proposed metalens has a full FOV of nearly 140 deg and achieves a diffraction-limited resolution of up to 1.55 μm at the center of the image side. The volume of neural meta-camera is 9.07 mm × 9.07 mm × 1.57 mm, which is integrated based on precision assembly platform.

Based on this meta-camera, we propose the knowledge-fused data-driven (KD) paradigm to address the image degradation problem. The KD paradigm is characterized by first initializing the transformer-based neural network using unsupervised PSF estimation, and then further fine-tuning the neural network using the data obtained from the meta-camera. In this way, a customized neural network can be trained to recover a range of imaging quality problems for the ultra-wide FOV metalens. The experiments on simple, cartoon, and complex scene images validate that our method solves the chromatic aberration, distortion, and central bright speckle of the meta-camera. Our work shows that the neural meta-camera can achieve ultra-wide FOV and full-color imaging, which is also difficult to obtain with conventional complex cameras.

2 Methods

2.1 On-Chip Neural Meta-Camera Model

Here, we demonstrate a miniature neural meta-camera for ultra-wide FOV and full-color imaging supported by a transformer-based image recovery neural network (Fig. 1). The network has a typical multiscale attention architecture and is trained under the guidance of the KD paradigm so as to improve the reconstructed image quality. As identified by yellow arrows in Fig. 1, the paradigm includes prior knowledge from simulated PSFs and data-driven measurements from the meta-camera, incorporating prior and measured dataset to initialize and fine-tune the network. On the other hand, the processing flow of the image recovery neural network follows the green arrows in Fig. 1. The images captured from the ultra-wide FOV meta-camera are reconstructed into ground-truth-like full-color images by the recovery neural network. With the help of the computility of the graphics processing units (GPUs), the model can conveniently repair the chromatic aberration, distortion, stray speckles, and background noise of the meta-camera.

2.2 Design Principle of the Ultra-Wide FOV Metalens

Recently, some approaches have been proposed for aberration correction and fast design of metasurfaces, such as hyperbolic phase profile,[12–15] quadratic phase optimization based on ray tracing,[25–27] gradient-based local optimization,[21] inverse design,[41–43] and combination of deep neural networks.[33–37,41–44] Here, in order to obtain ultra-wide FOV and accurate off-axis aberration correction on a CMOS image sensor plane [Fig. 2(a)], the phase profile of a 140 deg wide FOV metalens is optimized by the ray-tracing method.[25–27] Such a metalens is composed of a 220-μm-diameter aperture and a 1.54-mm-diameter metasurface that are located on both sides of a 0.7-mm-thick fused silica substrate, with an effective numerical aperture of 0.167 and an operating wavelength of 532 nm. The fact that the root mean square spot diagrams [right of Fig. 2(a)] on the sensor plane at different angles of incidence are all within the radius of Airy disks indicates the metalens's diffraction-limited performance with negligible monochromatic aberrations. We further simulate the alphabet image to illustrate good imaging performance in the whole FOV with clearly distinguishable alphabet letters [Fig. S1(a) in the Supplemental Material].

The metasurface contains Si nanoposts with different diameters arranged in quadrilaterals and covered by a 1-μm-thickness silicon dioxide protective layer. The phase coverage of 2π can be well achieved in seven selected nanoposts, with the average of over 95% transmission at normal incidence and decreasing value in off-normal directions. Note that the phase will shift accordingly when the incidence light is oblique. See more details in angle-dependent phase and transmission maps by rigorous coupled wave analysis[45] in Figs. S1(b) and S1(c) in the Supplemental Material. We emphasize the fact that the simulated modulation transfer function (MTF) curves of the metalens at different incident angles are very close to the diffraction limit case [Fig. 2(b)], demonstrating the effectiveness of the metalens for aberration correction over a wide full FOV. At different incidence angles, the simulated focusing efficiency of the metalens is 31.5% to 66.25% and decreases with the increase of incident angle due to the phase shift and nonuniform transmittance of nanoposts as the incident angle changes [Figs. S1(b) and S1(c) in the Supplemental Material].

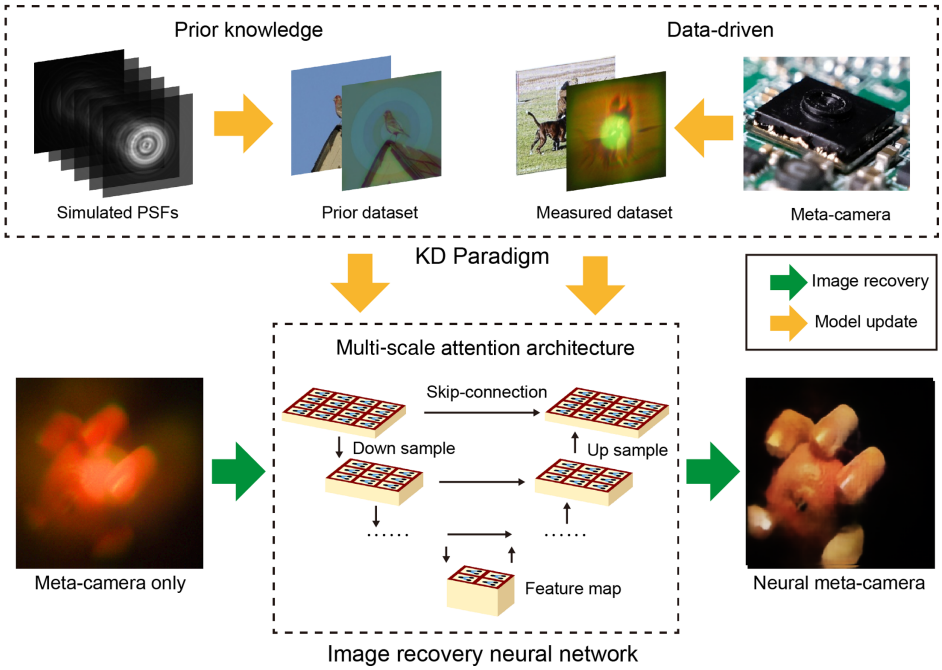

Fig. 1 Neural meta-camera model. The meta-camera consists of the ultra-wide FOV metalens and the transformer-based neural network for full-color imaging. Green arrows show the process of image recovery. The captured image from the meta-camera is reconstructed by the image recovery neural network constructed by the KD paradigm (yellow arrows, prior knowledge and data-driven). The neural network is initialized by the prior data set from the simulated PSFs of the metalens, and then the measured data set from meta-camera is input to drive the refinement of the initialized neural network. To capture information at multiscale, we use U-shaped hierarchical neural networks. Considering the spatial distribution characteristics of the simulated PSFs from the metalens, the U-shaped network with an attention mechanism is adopted to cope with its nonuniformity.

2.3 Demonstration of the Ultra-Wide FOV Metalens

The ultra-wide FOV metalens is fabricated by electron beam lithography and inductively coupled plasma-chemical vapor deposition. The aperture and metasurface are aligned through alignment marks patterned on both sides of a substrate (Fig. S2 in the Supplemental Material). Top-view scanning electron microscope (SEM) images of the fabricated metasurface highlight the excellent fabrication quality [Fig. 2(c)].

To evaluate the optical performance of the naked ultra-wide FOV metalens sample, we used an experimental setup that enables the metalens to focus a collimated light from different angles, and the focused spots to go into a rear microscopic system [Fig. S3(a) in the Supplemental Material]. One can see from Fig. 2(d) that the measured focal lengths (blue solid box) and the image heights (red solid box) are close to the simulations (dotted lines) from 0 deg to 70 deg at a center wavelength of 532 ± 5 nm. Note that the image height is defined as the offset position of focal PSFs from the optical axis center in the focal plane. The results show the capability of the metalens for a full FOV angular position, ensuring the accurate match between the metalens imaging plane and the CMOS image sensor. In addition, we compare the simulated and measured focal spots, full width at half-maximum values, and corresponding MTF curves of different incidence angles. More details can be found in Fig. S3 in the Supplemental Material.

To characterize the imaging resolution capability of the designed metalens, we use the measurement configuration shown in Fig. S4(a) in the Supplemental Material. The USAF 1951 resolution test chart is illuminated by the lamp with different narrowband filters, and the images can be captured by the microscopic system, including an objective lens, an adapter tube lens, and a CMOS sensor. The resolution test chart is fixed on the image plane, and the microscopic system moves along the optical axis to make the image clear. Figure 2(e) shows the projected images of the USAF 1951 resolution test chart at the angle of 0 deg and a center wavelength of 532 nm. The linewidth and gap in the vertical lines (yellow) and horizontal lines (orange) of element 3 in group 8 are clearly distinguished, and the corresponding contrast values are 35.9% and 37.5%, respectively [right side of Fig. 2(e)]. The contrast value is the ratio of the difference and sum of the maximum and minimum intensities. The contrast values are all above 20%, indicating that the resolution of the metalens in the center is 1.55 μm close to the diffraction-limited resolution ($\lambda/2$NA). The resolution results at wavelengths ranging from 488 to 680 nm are also shown in Fig. S4(b) in the Supplemental Material. We observe that the central field resolution of the ultra-wide FOV metalens is close to the diffraction limit at a single wavelength in the visible band.

To further characterize the wide FOV imaging capability, we select the number "7" of the USAF 1951 resolution test chart for

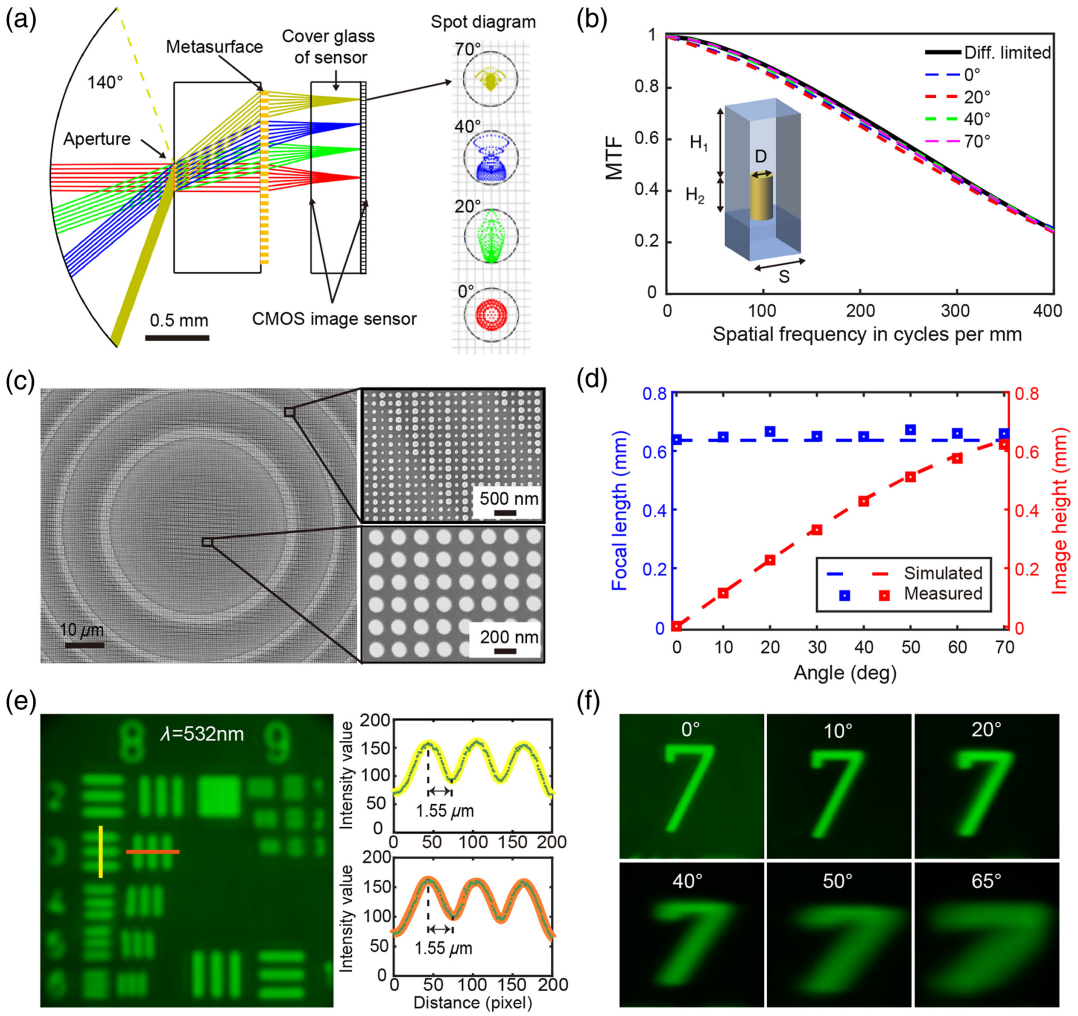

Fig. 2 Ray optics design and characterization of the ultra-wide FOV metalens. (a) Ray-tracing simulation results of ultra-wide FOV metalens (left) of 140 deg. The red/green/blue/yellow rays have four crossing points at the same image plane passing through the aperture, substrate, metasurface, and cover glass of sensor. Spot diagrams (right) show the diffuse spots with the incident angles of 0 deg, 20 deg, 40 deg, and 70 deg are located inside the Airy circle (black solid). (b) Simulated MTF curves at different FOVs and the black solid line indicate the diffraction limit. Schematic of a meta-atom of the metasurface, consisting of a silicon nanopost with the height (H_1) of 265 nm and a silicon dioxide protective layer with thickness (H_2) of 1 μm on a silica substrate. The nanoposts with varying diameters (D) are arranged in a square lattice with the lattice constant (S) of 220 nm. (c) Top-view SEM images of the metalens with different scales. (d) Simulated and measured focal length and image height of spots at different FOVs. (e) Projected images of the USAF 1951 resolution test chart at wavelengths of 532 nm. The corresponding intensity distributions of vertical lines (yellow) and horizontal lines (orange) of the element 3 from group 8 displayed a linewidth of 1.55 μm. (f) Image of the number 7 in different incident angles at the wavelength of 532 nm.

imaging. By changing the filters and turning the rotary stage, the images with projection angles from 0 deg to 70 deg can be captured at different wavelengths. When the angle of the rotary stage is 65 deg, the projected image of the number 7 reflects the angle range of about 63 deg to 70 deg. Figure 2(f) shows the projected images of the number 7 with projection angles of 0 deg, 10 deg, 20 deg, 30 deg, 40 deg, 50 deg, and 65 deg at the wavelength of 532 nm. The contours of the number 7 can be easily identified in the projected images at all angles, confirming the wide FOV imaging performance of the metalens. Additional experimental images of the number 7 at other wavelengths are shown in Fig. S4(c) in the Supplemental Material. Note that the distorted image with a projection angle greater than 40 deg is the inherent distortion of all wide FOV imaging systems, and it can be corrected by mature algorithms. As a result, the wide FOV imaging ability of the ultra-wide FOV metalens is confirmed by clearly demonstrating the projection imaging in the range of 0 deg to 70 deg half-FOV.

2.4 KD Paradigm with Transformer-Based Network

Due to its self-attention mechanism design, the transformer module can capture longer distance context relationships, which can be interpreted as a global relationship modeling for image processing tasks.[46] In the design of the ultra-wide FOV metalens at single-wavelength, PSFs of other wavelengths often suffer from severe mass loss, manifesting in the form of unconcentrated energy distribution, unfocused diffuse spots (Fig. S5 in the Supplemental Material), etc. These problems make the modeling of ultra-wide FOV metalens imaging more difficult for neural networks, and previous work has used traditional neural network architectures;[47] however, the existing methods are still struggling to deal with such complex degradations. Fortunately, the transformer-based networks can handle the complex degradation described above for the ability of modeling long-distance dependencies.

In addition to the network structure, we point out that the training paradigm is also crucial. Considering the incompleteness of the theoretical simulation of the imaging process and the difference between theory and actual fabrication, the distortion and central bright speckle of the ultra-wide FOV metalens in the visible spectrum imaging hinder learning an effective model based on the pure theoretical approximation. Recent research has shown that deep-learning models trained at a large scale on similar tasks can learn transferable domain knowledge, so that it can be adapted to downstream tasks by a transfer learning manner.[48] Therefore, we propose a two-stage paradigm to train a transformer network to recover the chromatic aberrations, distortion, and central bright speckle in the metalens imaging.

Figure 3 shows the proposed KD paradigm, including two stages: prior knowledge and data-driven. In the first stage shown in Fig. 3(a), we leverage the prior knowledge of metalens design to initialize the model with design parameters of the metalens in an unsupervised manner. Then, we perform data-driven learning to refine our neural network based on the collected real data in the second stage shown in Fig. 3(b) to drive its performance close to a conventional commercial lens. We use the same attention-based U-structured neural network[49] (right part of Fig. 3) in both stages, so we can extract multiscale features and ensure that the recovered images of metalens are semantically consistent at various scales, producing a high-quality image, as expected. Note that we use the same loss function based on mean squared error in both stages as well.

Specifically, we first use the theoretical design parameters of the metalens and the theory of angular spectral propagation to simulate the PSF sets of the metalens in different FOVs and wavelengths.[33] Since the design of the metalens is circularly symmetric, it is convenient to rotate these PSFs to obtain approximate PSFs of full fields collection,

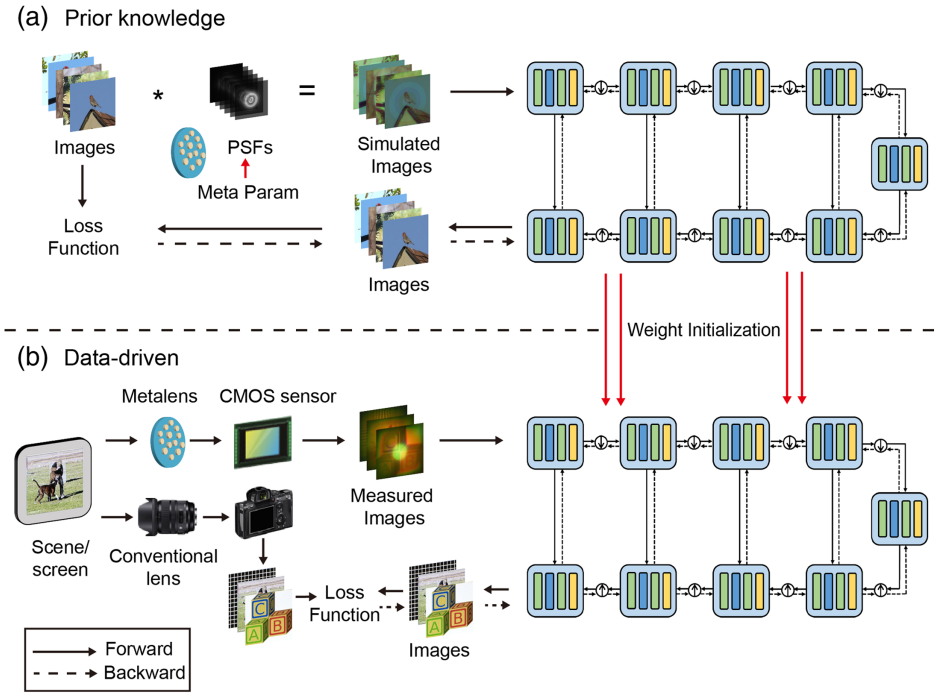

Fig. 3 Proposed KD paradigm for training image recovery neural network. (a) Prior knowledge, i.e., PSFs, obtained from the design parameters of the metalens is applied to the original images to generate the prior data set. This prior data set is used to train an initialized neural network. (b) By utilizing the data collection and processing flow we have established, data from corresponding scenarios are collected to drive further fine-tuning of the model, enabling it to cope with more intricate image degradation in actual scenarios. The measured data set in the data-driven scenario includes images (e.g., LCD screen projection images) captured by the metalens and a conventional commercial lens (Sigma Art Zoom lens). As shown by the black dotted line, the neural network is updated through backpropagation with the same loss function in both stage (a) and stage (b). After the model parameters updates of two stages, the neural network is employed to recover imaging in the corresponding scenario.

$$\text{Image}(\lambda)_{\text{meta}} = \sum_{\theta}\sum_{\phi} \text{mask}(\theta,\phi,\lambda)\cdot [\text{psf}(\theta,\phi,\lambda) \otimes \text{Image}(\lambda)_H], \quad (1)$$

where $\text{Image}(\lambda)_{\text{meta}}$ means simulated image channel corresponding to wavelength λ, $\text{mask}(\theta,\phi,\lambda)$ and $\text{psf}(\theta,\phi,\lambda)$ are the mask and PSF in theory corresponding to FOV θ, rotation angle ϕ, and wavelength λ, respectively, and $\text{Image}(\lambda)_H$ represents an image corresponding to wavelength λ to be convolved. It is worth noting that, compared to previous works that convolve images with PSF of a single incident angle at each wavelength,[34] we incorporate the PSF of all incident angles into the simulation, so that we are able to take into consideration the strong variations of the PSF at nondesignated wavelengths during our modeling. The detailed processing about PSF generation can be found in Section S3 in the Supplemental Material. Note that the data set we collected includes the aberration information of each FOV, so that our initialized neural network can capture the prior knowledge about the aberration distribution of the imaging, allowing the network to achieve faster convergence and better performance in the second stage.

We use the data-driven approach instead of the measured PSF set-driven method[33,34] in the second stage to circumvent the following problems. Existing single-wavelength wide FOV metalenses with a small front aperture have a central bright speckle problem at nondesigned wavelengths, which becomes serious with the increase of the incident angle. Unfortunately, so far there are no accurate theoretical models to estimate the central bright speckle. Moreover, the intensity variation and spatial inhomogeneity of PSFs at different angles of incidence and at nondesigned wavelengths make it difficult for the measured PSF set to restore the real image effect. With such large differences in PSF intensities, the measured PSF set ensemble will have a greater loss of precision, resulting in a more tedious and arduous task to measure the PSF set than our data-driven method.

In the second stage [Fig. 3(b)], we build an image acquisition processing system to efficiently acquire real data for fine-tuning our model. The image acquisition processing system shots the images displayed on the LCD screen (Portkeys LH5P II, 5.5″, 1920×1080) as the scenes by a conventional commercial lens (Sigma Art Zoom lens) or metalens, and finally collects the image pair captured by the CMOS sensor (e.g., IMAX 335) and commercial Sony sensor (e.g., A7M3, Sony), respectively. More details about this image acquisition processing system can be found in Section S4 in the Supplemental Material. Section S5 in the Supplemental Material further describes our data processing procedures; that is, once the process is established, it may be possible to cascade data-processing flows and neural networks to quickly process imaging. The ablation experiments shown in Section S7 in the Supplemental Material demonstrate the effectiveness of our method.

In addition, we enhance the model by using the equivariant in imaging process throughout the experiment by the following equation:

$$T(I) = T(\text{psf}\cdot I), \quad (2)$$

where T is a particular transformation, I is the imaged object, and psf is the PSF corresponding to the one-to-one imaging process. By utilizing the equivariant of physical processes to augment data, the model can discover potential physical properties for better robustness on unseen data.[50]

3 Results

3.1 Naked Metalens for Neural Imaging

To demonstrate the performance of the ultra-wide FOV metalens combined with the neural network, we conduct an experimental comparison by imaging different types of images in the image acquisition processing system. Considering the trade-off between data collection cost and recovery effectiveness, we collected 1000 images to validate our approach, 800 as training data and 200 as test data. As shown in Fig. 4(a), the image data of scenes (e.g., projected by the LCD screen) are imaged by the naked ultra-wide metalens, and then captured by the microscopic system consisting of a 10× objective (MPLFLN10×BD, Olympus), an adapter tube lens (1-62922, Navitar), and a CMOS sensor (A7M3, Sony). Original images captured by the metalens and corresponding recovery results from our neural networks, UNet & KD paradigm (UNet trained with KD paradigm), and other traditional image enhancement algorithms are shown in Fig. 4(b). Compared with the unrecovered image of the naked ultra-wide FOV metalens on the leftmost of Fig. 4(b), the contrast and sharpness of the images restored by the sharpened Laplacian algorithm and the multiscale retinex with color restoration (MSRCR) algorithm are not improved much, due to uncorrected background noises. The images recovered by the UNet & KD paradigm can effectively eliminate the central bright speckle, but the contrast and sharpness of the images are not good enough. In contrast, high-contrast and panchromatic aberration correction images can be recovered by our method (transform-based neural network trained with KD paradigm). From the zoom-in images in Fig. 4(b), it is clear that the contrast of the object's contour boundaries has been well refined, and the contour boundaries no longer have color overlay vignetting due to magnification chromatic aberration. More information on the comparison of other traditional convolutional networks with our image recovery neural network (transform-based network) is provided in Section S8 in the Supplemental Material. Therefore, our image recovery neural network offers a considerable enhancement in color similarity, contrast, and edge sharpness compared to traditional algorithms and other traditional convolutional networks.

3.2 Meta-Camera for Neural Imaging

To demonstrate a proof-of-concept application, we package the metalens with a CMOS image sensor into a miniature and portable meta-camera with a volume of 9.07 mm × 9.07 mm × 1.57 mm. Figure 5(a) shows the photograph of the meta-camera system, including the diaphragm, sleeve, base, CMOS image sensor (IMX335, Sony), and core optical element of the wide FOV metalens. The advancement of our proposed compact integration approach is that we have built a precision assembly platform to ensure the integrated camera modules are versatile and practical. The professional design of the support structure greatly reduces the complexity and difficulties caused by inclination and eccentricity in assembly. The most critical step in the assembly process is to ensure that the distance between the metalens sample and the CMOS sensor is accurate enough. For this purpose, the thread structure is designed and manufactured between the sleeve and the base to facilitate precise adjustment of the image clarity of

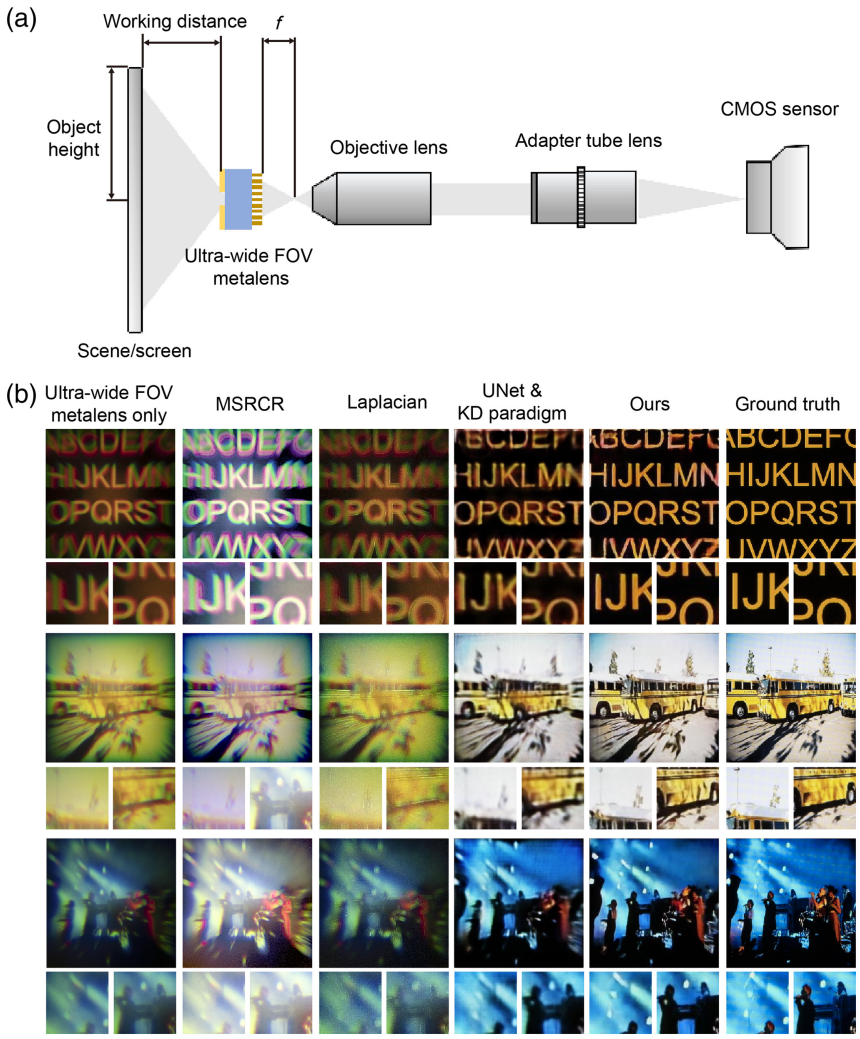

Fig. 4 Image recovery results of our neural network for images of naked ultra-wide FOV metalens are compared with results from UNet & KD paradigm and other traditional methods. (a) Schematic illustration of the data acquisition system for naked ultra-wide FOV metalens. The object projected by a 5.5-in. LCD screen is collected by the naked ultra-wide FOV metalens with a working distance of 2 cm and redirected to a micro-magnification system with an objective lens (Olympus, MPLFLN10xBD), an adapter tube lens (1-62922, Navitar), and a CMOS sensor (Sony, A7M3). (b) Compared to UNet & KD paradigm and other traditional image recovery algorithms (e.g., MSRCR, Laplacian), our image recovery neural network produces ultra-wide FOV, full-color and high-quality images corrected for central bright speckle, chromatic aberrations, and distortion. Examples of recovered images include complex scenes, such as cartoons with orange alphabets, yellow buses in the shade, and concerts under blue lights. Detail insets are illustrated below each row. Compared to ground-truth capture (the rightmost column) using a conventional commercial lens (Sigma Art 24-70mm DG DN), our neural network accurately reproduces fine details and colors in images. More comparison images (e.g., grids, letters, and oranges) are shown in Figs. S12–S14 in the Supplemental Material.

the camera module. In addition, to ensure an accurate bond between the components, we use ultraviolet curing adhesive for sealing with a curing time of 2 min.

To exhibit the capability of the neural meta-camera, we placed an LCD screen at different working distances from the meta-camera so that it could capture images with a large FOV [Fig. 5(b)]. Following the setting in the metalens demonstration, we use 800 images for training and 200 images for evaluation. Figure 5(c) shows the results at a working distance of 2 cm before and after recovery of the neural meta-camera and the neural ultra-wide FOV metalens. Compared to the ultra-wide FOV metalens, the original images captured by the meta-camera have a more severe central bright speckle and color cast. The exacerbation of the central bright speckle is due to the burr and irregular shape of the aperture of the diaphragm caused by a processing error, while the color cast is derived from the difference in spectral response curve of CMOS image sensors between commercial Sony sensor (A7M3) and IMX335.

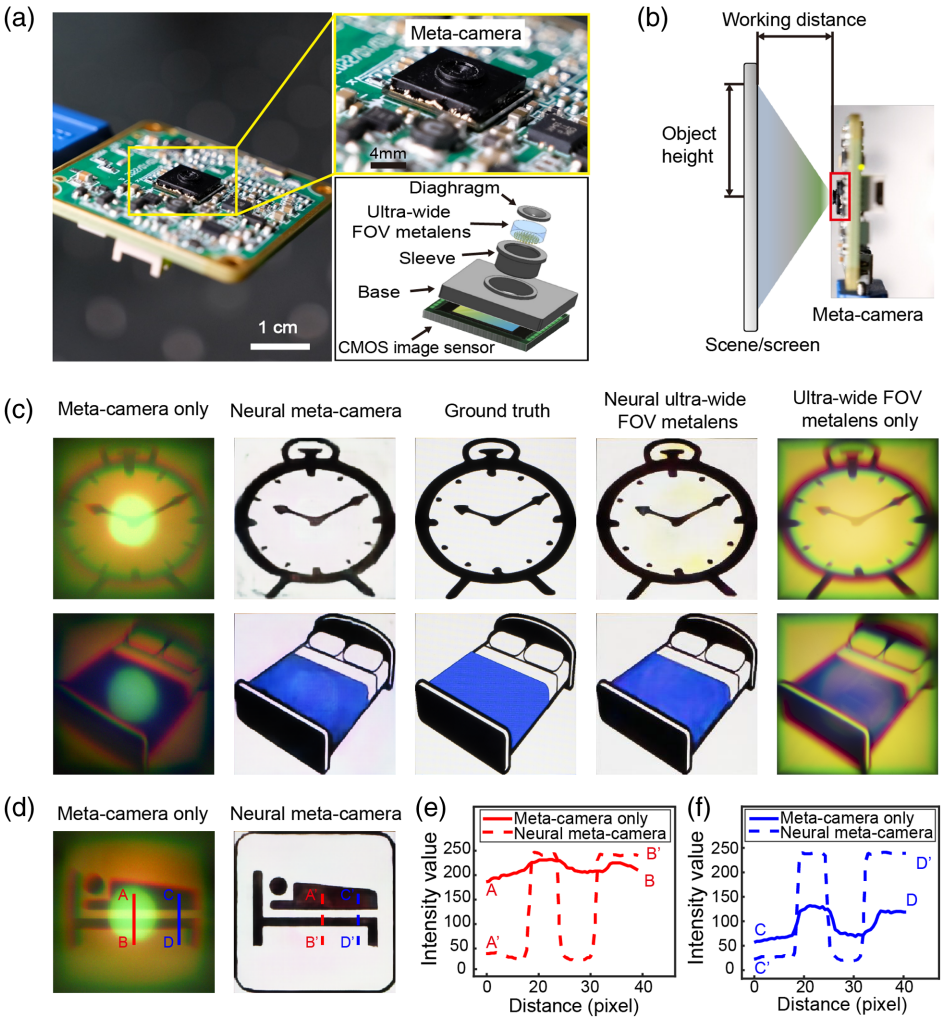

Fig. 5 Neural meta-camera for imaging. (a) Photograph of the meta-camera system (left) by integrating the miniature meta-camera (top-right) with a CMOS image sensor, and the schematic illustration of its structural mechanism (bottom-right) including an aperture, sleeve, and base for shading and waterproofing. (b) Schematic diagram of meta-camera test. The ground-truth images are projected on the LCD screen and captured directly by the meta-camera. (c) Comparison recovery results from images captured by ultra-wide FOV metalens only and the meta-camera at the working distance of 2 cm. Cartoon images from an alarm clock and a blue bed show that chromatic aberrations and central bright speckle are greatly improved after recovery by neural networks. More comparison images (e.g., doll, coral, and concert) are shown in Figs. S15–S16 in the Supplemental Material. (d) Images captured through the meta-camera only or with the neural meta-camera. (e), (f) The corresponding intensity profiles along line AB, A'B', CD, and C'D' in the central and edge areas of the images, respectively. The image contrast for the neural meta-camera exhibits substantial enhancement compared to that for the meta-camera without neural networks.

Cartoon images from an alarm clock and a blue bed show that chromatic aberrations and central bright speckle are greatly improved after recovery through our method. The attention mechanism leads to a wider receptive field, combined with a multiscale structure, allowing for a more complete removal of global information-related bright speckle in a central position. Despite the images captured by the meta-camera having a stronger bright speckle than those captured by the ultra-wide FOV metalens only, the proposed neural network can still eliminate them. To quantitatively evaluate the performance of the neural meta-camera, we test a black-and-white target image. The captured images of the black-and-white target are shown in Fig. 5(d); the image from the neural meta-camera has no central bright speckle and color casts, and the line contours are clearer than those using only the meta-camera. Figures 5(e) and 5(f) show the intensity distribution at the center and edge of the captured image, where the solid and dashed lines correspond to images captured only from the meta-camera and improved by the neural network, respectively. The calculated contrasts at the center and edge parts of the target images are increased by 13.5 times and 2.7 times, i.e., 0.834, 0.846 for the neural meta-camera, and 0.062, 0.313 for the meta-camera only, respectively. The high contrast values

indicate high edge sharpness in neural meta-camera imaging. In conclusion, our neural meta-camera enables high-quality, wide FOV, and full-color imaging.

To assess the practicability and feasibility of the neural meta-camera in an actual scene, we captured and recovered the images in two scenarios. One is the imaging of three monitor screens at different working distances; the other is the imaging of multiple objects of various colors arranged at different depths in an actual scene. In the first scenario, we obtained recovery images at the working distances of 1.3, 12, and 44.5 cm, as shown in Fig. S17 in the Supplemental Material. It can be seen that the image restoration clarity and color comparison are uniform at different working distances. The calculated peak signal-to-noise ratio and structure similarity index measure (SSIM) values (as shown in Table S4 in the Supplemental Material) further emphasize quantitatively the quality of image restoration at different distances.

In the other scenario, we further capture and recover the image of letters and dolls at different working distances in an indoor scene. We set up a dual optical path data acquisition system [Fig. S18(a) in the Supplemental Material] based on a cube beam splitter to obtain pixel-level aligned data sets. As shown in Fig. S18(b) in the Supplemental Material, in the recovered image from the neural meta-camera, the letters are clearer, and the dolls at different working distances of 40, 55, and 85 cm can also be identified. Although the recovered image lacks detail, its central bright speckle and chromatic aberration are greatly improved compared to the original image from the meta-camera. In addition, based on the imaging data from the actual scene, we compare the performance between the imaging of the meta-camera and the traditional camera on the multi-label image classification task. The data from the meta-camera achieve a precision of 96.47%, while the data from traditional camera achieve 96.73%. Experiments demonstrated that the imaging of the meta-camera did not show significant performance differences in recognition tasks compared to imaging from a traditional camera, which hints at the potential of meta-camera for classification and recognition application.

4 Discussion and Conclusion

Our work demonstrates a neural meta-camera for ultra-wide FOV and full-color imaging in single shot without scanning or image stitching. The neural meta-camera consists of an ultra-wide FOV metalens, a CMOS image sensor, and the image recovery neural network. Due to the high-precision assembly technology, our neural meta-camera is only 9.07 mm × 9.07 mm × 1.57 mm in volume, including the support structure and the CMOS image sensor. The neural meta-camera overcomes chromatic aberration, distortion, central bright speckle, and background noise through image recovery neural network and successfully achieves full-color imaging with high contrast over a wide FOV. Such a neural meta-camera is an exemplary case in imaging systems with miniaturization, functionality, wide FOV, and high-quality performance at the same time.

The proposed KD paradigm is theoretically uncoupled from the design approach, so it is extended to applications such as depth of field synthesis and outdoor imaging. Under ideal conditions, the model can recover images at the speed of 48 frames per second on the RTX 3090 GPU, which opens up the possibility of real-time[51] processing in the future. This novel neural meta-camera module paves the route for meta-optics for the thinner, lightweight, and more compact visible full-color imaging system, such as noninvasive[52] endoscopy, robot navigation, micro-intelligent systems, and engineering surveying.

Code and Data availability

The data and code supporting this study are available from the corresponding authors upon reasonable request.

Acknowledgments

This work was supported by the National Key R&D Program of China (Grant No. 2023YFB2806800), the National Natural Science Foundation of China (Grant Nos. 62035016 and U21A20471), the Guangdong Basic and Applied Basic Research Foundation (Grant No. 2023B1515040023), and the Guangzhou Science and Technology Program (Grant No. 202201011671).

References

1. J. Wu et al., "An integrated imaging sensor for aberration-corrected 3D photography," *Nature* **612**, 62–71 (2022).
2. K. Kim et al., "Biologically inspired ultrathin arrayed camera for high-contrast and high-resolution imaging," *Light Sci. Appl.* **9**, 28 (2020).
3. Z.-Y. Hu et al., "Miniature optoelectronic compound eye camera," *Nat. Commun.* **13**, 5634 (2022).
4. Y. Zhou et al., "Flat optics for image differentiation," *Nat. Photonics* **14**, 316–323 (2020).
5. M. K. Chen et al., "Principles, functions, and applications of optical meta-lens," *Adv. Opt. Mater.* **9**, 2001414 (2021).
6. A. Arbabi and A. Faraon, "Advances in optical metalenses," *Nat. Photonics* **17**, 16–25 (2023).
7. M. Pan et al., "Dielectric metalens for miniaturized imaging systems: progress and challenges," *Light Sci. Appl.* **11**, 195–226 (2022).
8. B. B. Xu et al., "Metalens-integrated compact imaging devices for wide-field microscopy," *Adv. Photonics* **2**, 066004 (2020).
9. X. Luo et al., "Recent advances of wide-angle metalenses: principle, design, and applications," *Nanophotonics* **11**, 1–20 (2022).
10. F. Yang et al., "Wide field-of-view metalens: a tutorial," *Adv. Photonics* **5**, 033001 (2023).
11. A. Arbabi et al., "Dielectric metasurfaces for complete control of phase and polarization with subwavelength spatial resolution and high transmission," *Nat. Nanotechnol.* **10**, 937–943 (2015).
12. A. Arbabi et al., "Subwavelength-thick lenses with high numerical apertures and large efficiency based on high-contrast transmit arrays," *Nat. Commun.* **6**, 7069 (2015).
13. M. Khorasaninejad et al., "Metalenses at visible wavelengths: diffraction-limited focusing and subwavelength resolution imaging," *Science* **352**, 1190 (2016).
14. Z.-B. Fan et al., "Silicon nitride metalenses for close-to-one numerical aperture and wide-angle visible imaging," *Phys. Rev. Appl.* **10**, 014005 (2018).
15. H. Liang et al., "Ultrahigh numerical aperture metalens at visible wavelengths," *Nano Lett.* **18**, 4460–4466 (2018).
16. S. Shrestha et al., "Broadband achromatic dielectric metalenses," *Light Sci. Appl.* **7**, 85 (2018).
17. S. Wang et al., "A broadband achromatic metalens in the visible," *Nat. Nanotechnol.* **13**, 227–232 (2018).
18. Z.-B. Fan et al., "A broadband achromatic metalens array for integral imaging in the visible," *Light Sci. Appl.* **8**, 67 (2019).
19. R. J. Lin et al., "Achromatic metalens array for full-colour light-field imaging," *Nat. Nanotechnol.* **14**, 227–231 (2019).
20. H. Li et al., "Bandpass-filter-integrated multiwavelength achromatic metalens," *Photonics Res.* **9**, 1384–1390 (2021).
21. Z. Li et al., "Meta-optics achieves RGB-achromatic focusing for virtual reality," *Sci. Adv.* **7**, eabe4458 (2021).

22. W. Feng et al., "RGB achromatic metalens doublet for digital imaging," *Nano Lett.* **22**, 3969–3975 (2022).
23. A. Arbabi et al., "Miniature optical planar camera based on a wide-angle metasurface doublet corrected for monochromatic aberrations," *Nat. Commun.* **7**, 13682 (2016).
24. B. Groever, W. T. Chen, and F. Capasso, "Meta-lens doublet in the visible region," *Nano Lett.* **17**, 4902–4907 (2017).
25. Y. Liu et al., "Meta-objective with sub-micrometer resolution for microendoscopes," *Photonics Res.* **9**, 106–115 (2021).
26. M. Y. Shalaginov et al., "Single-element diffraction-limited fisheye metalens," *Nano Lett.* **20**, 7429–7437 (2020)
27. F. Zhang et al., "Extreme-angle silicon infrared optics enabled by streamlined surfaces," *Adv. Mater.* **33**, 2008157 (2021).
28. A. Martins et al., "On metalenses with arbitrarily wide field of view," *ACS Photonics* **7**, 2073–2079 (2020).
29. J. Chen et al., "Planar wide-angle-imaging camera enabled by metalens array," *Optica* **9**, 431–437 (2022).
30. S. Colburn, A. Zhan, and A. Majumdar, "Metasurface optics for full-color computational imaging," *Sci. Adv.* **4**, eaar2114 (2018).
31. E. E. Fenimore, "Coded aperture imaging: the modulation transfer function for uniformly redundant arrays," *Appl. Opt.* **19**, 2465–2471 (1980).
32. S. R. Gottesman and E. E. Fenimore, "New family of binary arrays for coded aperture imaging," *Appl. Opt.* **28**, 4344–4352 (1989).
33. E. Tseng et al., "Neural nano-optics for high-quality thin lens imaging," *Nat. Commun.* **12**, 6493 (2021).
34. Q. Fan et al., "Trilobite-inspired neural nanophotonic light-field camera with extreme depth-of-field," *Nat. Commun.* **13**, 2130 (2022).
35. Y. Peng et al., "The diffractive achromat full spectrum computational imaging with diffractive optics," *ACM Trans. Graphics* **35**, 31 (2016).
36. V. Sitzmann et al., "End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging," *ACM Trans. Graphics* **37**, 114 (2018).
37. Y. Peng et al., "Learned large field-of-view imaging with thin-plate optics," *ACM Trans. Graphics* **38**, 219 (2019).
38. L. Yann, B. Yoshua, and H. Geoffrey, "Deep learning," *Nature* **521**, 436–444 (2015).
39. J. Jumper et al., "Highly accurate protein structure prediction with AlphaFold," *Nature* **596**, 583–589 (2021).
40. T. Brown et al., "Language models are few-shot learners," in *Adv. Neural Inf. Process. Syst. 33*, pp. 1877–1901 (2020).
41. S. Pinilla et al., "Miniature color camera via flat hybrid meta-optics," *Sci. Adv.* **9**, eadg7297 (2023).
42. S. Molesky et al., "Inverse design in nanophotonics," *Nat. Photonics* **12**, 659–670 (2018).
43. W. Ma et al., "Deep learning for the design of photonic structures," *Nat. Photonics* **15**, 77–90 (2021).
44. F. Wang et al., "Phase imaging with an untrained neural network," *Light Sci. Appl.* **9**, 77 (2020).
45. V. Liu and S. Fan, "S^4: a free electromagnetic solver for layered periodic structures," *Comput. Phys. Commun.* **183**, 2233–2244 (2012).
46. A. Vaswani et al., "Attention is all you need," in *Proc. 31st Annu. Conf. Neural Inf. Process. Syst. (NIPS)*, pp. 6000–6010 (2017).
47. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pp. 234–241 (2015).
48. C. Tan et al., "A survey on deep transfer learning," in *Int. Conf. Artif. Neural Networks (ICANN)*, pp. 270–279 (2018).
49. Z. Wang et al., "Uformer: a general u-shaped transformer for image restoration," in *Proc. IEEE/CVF Conf. Comput. Vision and Pattern Recognit. (CVPR)*, pp. 17683–17693 (2022).
50. D. Chen, J. A. Tachella, and M. E. Davies, "Robust equivariant imaging: a fully unsupervised framework for learning to image from noisy and partial measurements," in *Proc. IEEE/CVF Conf. Comput. Vision and Pattern Recognit. (CVPR)*, pp. 5647–5656 (2022).
51. J. Xiong et al., "Dynamic brain spectrum acquired by a real-time ultraspectral imaging chip with reconfigurable metasurfaces," *Optica* **9**, 461–468 (2022).
52. L. Zhu et al., "Large field-of-view non-invasive imaging through scattering layers using fluctuating random illumination," *Nat. Commun.* **13**, 1447 (2022).
53. Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.* **22**, 1330–1334 (2000).
54. D. J. Jobson, Z. U. Rahman, and G. A. Woodell, "A multiscale retinex for bridging the gap between color images and the human observation of scenes," *IEEE Trans. Image Process.* **6**, 965–976 (1997).

Yan Liu received her PhD from the School of Physics, Sun Yat-sen University, China, in 2023. She is currently a postdoctoral fellow at the School of Physics of Sun Yat-sen University. Her current research interests include meta-optics and meta-device for imaging, sensing, and display.

Wen-Dong Li received his BS degree from the School of Electronic Information, Sichuan University, China, in 2019. He is currently a PhD student at the School of Computer Science and Engineering, Sun Yat-sen University. His research interests focus on computer vision, especially image reconstruction and 3D vision.

Kun-Yuan Xin received her BE degree from the College of Science and Engineering, Jinan University, China, in 2021. She is currently a PhD student at the School of Physics, Sun Yat-sen University, China. Her research interests focus on metasurfaces imaging.

Wei-Shi Zheng is a full professor with Sun Yat-sen University and is working on AI, especially focusing on video and image understanding. He has published more than 200 papers. He is an associate editor on the editorial board of IEEE TPAMI and IEEE TAI journals. He is a Cheung Kong Scholar Distinguished Professor, a recipient of the NSFC Excellent Young Scientists Fund, and a recipient of the Royal Society-Newton Advanced Fellowship of the United Kingdom.

Jian-Wen Dong is a full professor in the School of Physics at Sun Yat-sen University. He has been studying metaphotonics and subwavelength optical structures, including (1) topological physics/photonics and (2) metasurface for computational imaging and 3D display. He is a Cheung Kong Scholar Young Professor. and a recipient of the NSFC Excellent Young Scientists Fund. He was awarded the prizes of the Youth Science Award of the Ministry of Education and the Top Ten China Optics Award.

Biographies of the other authors are not available.