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

Yan Liu[†1], Wen-Dong Li[†2], Kun-Yuan Xin[†1], Ze-Ming Chen[1], Zun-Yi Chen[1], Rui Chen[1], Xiao-Dong Chen[1], Fu-Li Zhao[1], Wei-Shi Zheng[*2], and Jian-Wen Dong[*1]

[1]State Key Laboratory of Optoelectronic Materials and Technologies & School of Physics, Sun Yat-sen University, Guangzhou 510275, China
[2]School of Computer Science and Engineering, Sun Yat-sen University, 510006, Guangzhou, China

[†] These authors contributed equally to this work.
[*] Corresponding authors: Wei-Shi Zheng, E-mail: zhwshi@mail.sysu.edu.cn ; Jian-Wen Dong, Email: dongjwen@mail.sysu.edu.cn

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 ultra-thin optical components, but there are a set of tricky challenges like chromatic aberrations correction, central bright speckle removal, and image quality improvement of wide FOVs. In this paper, we design a neural meta-camera by introducing a knowledge-fused data-driven (KD) paradigm equipped with transformer-based network. Such 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 distance validate that the proposed meta-camera can

achieve ultra-wide FOV (> 100-degree) and full-color image 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 capabilities, the neural meta-camera emerges as a compelling alternative for applications such as micro-navigation, micro-endoscopes, and various on-chip devices.

1 Introduction

Conventional cameras are renowned for their large imaging field of view 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, lightweight and portability cameras[1-3] are increasingly desired in autonomous driving, endoscopic medical, 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 a potential alternatives to bulky and complex optical instruments[4-9]. Metalens exhibits superior optical performance due to its ability to precisely manipulate the incidence beam;[10-15] however, it still remains challenge of aberration correction, in particular for chromatic aberration and off-axis monochromatic aberration. To eliminate chromatic aberration, dispersive propagation phase and dispersive-free geometic phase have been introduced to achieve broadband[16-19] and multi-wavelength[20-22] achromatic metalenses. Due to the limitation of the group delay dispersion of meta-atoms, achromatic metalenses are usually implemented on paraxial. In the off-axis case, a lot of efforts have been made to correct the off-axis monochromatic aberrations of the metalens. By introducing ray tracing method,[23-27] Fourier analysis[28] and metalens array[29], cascaded metalens located on either side of the substrate[23-25] and single wide FOV metalens with an aperture[26,27] can correct off-axis monochromatic aberrations and achieve diffraction-limited imaging over wide FOV. However, chromatic aberration correction, central bright speckle, and image quality improvement of wide FOV metalens in the off-axis are rarely considered. Obviously, addressing the aforementioned issues of wide FOV metalens relies on the above existing methods remains a considerable challenge.

To improve the image quality of metalens, traditional image restoration computational imaging methods[2,22,30,51,52] are introduced, and they usually recover images based on simple hypothesis or enhance images through multiple-image super-resolution.[2,22,30] However, the factors influencing the imaging quality of current ultra-wide FOV metalens are intricate, making it difficult to improve the imaging quality based on a single hypothesis. Recent years, several methods are proposed to incorporate neural networks for improving the imaging quality of metalens or diffractive optical elements, which use point spread functions (PSFs) to train models.[31-35] Unfortunately, single-wavelength ultra-wide FOV metalens have complex PSFs spatial variation in different incident angles at other wavlengths, which makes it fail to accurately model the imaging degradation by applying the above PSFs method. Even worse, the central bright speckle indicates that there is inconsistence between the simulation data and the actual scene, which makes it more difficult to improve the imaging quality of ultra-wide FOV metalens by the prior PSFs method.

With the advancement of deep learning[36] research, transformer modules based on attention mechanisms have been developed and demonstrated to be effective in cutting-edge studies such as AlphaFold2,[37] GPT,[38] large image-text models, etc. 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 expectable that incorporating transformer methodology into wide FOV metalens imaging is a good choice to cope with more complex PSFs 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 ultra-wide FOV metalens assembled on a CMOS image sensor. The proposed metalens has a full FOV of nearly 140°, 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 $9.07 \times 9.07 \times 1.57$ mm^3, which is integrated based on precision assembly platform.

Based on this meta-camera, we propose the knowledge-fused data-driven (KD) paradigm to address image degradation problem. A characteristic of the proposed KD paradigm is to first initializes the transformer-based neural network using the PSF estimation in unsupervised manner, and then the data acquired from the meta-camera is used to further fine-tune the neural network. 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, distortions 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 transformer-based image recovery neural network (Fig. 1). The network is a typical multi-scale attention architecture and 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 meta-camera, incorporating prior and measured datasets 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 (GPU), the model can conveniently repair the chromatic aberration, distortion, stray speckles, and background noise of the meta-camera.

2.2 Design principle of ultra-wide FOV metalens

Recently, some approaches have been proposed for aberration correction and fast design of metasurface, such as hyperbolic phase profile,[12-15] quadratic phase optimization based on ray tracing,[25-27] gradient-based local optimization,[21] inverse design,[39-41] and combination of deep neural networks.[31-35,39-42] 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° wide FOV metalens is optimized by ray tracing method.[25-27] Such 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-in the radius of Airy disks, indicates the metalens' diffraction-limited performance with negligible monochromatic aberrations. We further simulate the alphabet image to illustrate good imaging performance in whole FOV with clearly distinguishable alphabet letters [Fig. S1(a)].

The metasurface contains Si nanoposts with different diameters arranged in quadrilaterals and covered by 1-μm-thickness silicon dioxide protective layer. The phase coverage of 2π can be well achieved in seven selected nanoposts, with the average over 95% transmission at normal incidence and decreasing value at off-normal. Note that the phase will shift accordingly when oblique as well. See more detail in angle-dependent phase and transmission maps by rigorous coupled wave analysis[44] in Figs. S1(b-c). 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 are 31.5%~66.25%, and decreases with the increase of incident angle due to the phase shift and non-uniform transmittance of nanoposts as the incident angle changes [Fig. S1(b-c)].

2.3 Demonstration of ultra-wide FOV metalens

The ultra-wide FOV metalens is fabricated by electron beam lithography (EBL) and inductively coupled plasma-chemical vapor deposition (ICP-CVD). The aperture and metasurface are aligned through alignment marks patterned on both sides of a substrate (Fig. S2). Top-view scanning electron microscope (SEM) images of the fabricated metasurface, highlighting 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 focusing a collimated light from different angles and the focused spots well-going into a rear microscopic system [Fig. S3(a)]. 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° to 70° 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 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 maximums (FWHM) values and corresponding MTF curves of different incidence angles. More details can be found in Fig. S3.

To characterize the imaging resolution capability of the designed metalens, we use the measurement configuration shown in Fig. S4(a). 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, 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. Fig. 2(e) shows the projected images of the USAF 1951 resolution test chart at the angle of 0° 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 (λ/2NA). The resolution results at wavelengths ranging from 488nm to 680nm are also shown in Fig. S4(b). 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 imaging. By changing the filters and turning the rotary stage, the images with projection angles from 0° to 70° can be captured at different wavelengths. When the angle of the rotary stage is 65°, the projected image of the number "7" reflects the angle range of about 63° to 70°. Fig. 2(f) shows the projected images of the number "7" with projection angles of 0°, 10°, 20°, 30°, 40°, 50°, and 65° at the wavelength of 532 nm. The contours of 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). Note that the distorted images with a projection angle greater than 40° 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 ultra-wide FOV metalens is confirmed by clearly demonstrating the projection imaging in the range of 0~70° half-FOV.

2.4 KD paradigm with transformer-based network

Due to its self-attention mechanism design, transformer module can capture longer distance context relationships, which can be interpreted as a global relationship modeling for image processing tasks.[44] In the design of ultra-wide FOV metalens at single-wavelength, PSFs of other wavelengths often suffer from severe mass loss, manifesting itself in the form of unconcentrated energy distribution, unfocused diffuse spots (Fig. S5), 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;[45] 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, it is pointed 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 visible spectrum imaging hinder learning an effective model based on the pure theoretical approximation. Recent research have shown that deep learning models trained at large-scale on similar tasks can learn transferable domain knowledge, so that it can be adapted to downstream tasks by transfer learning manner.[46] 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.

Fig. 3 illustrates the proposed KD paradigm, including two stages of 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 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 conventional commercial lens. We use the same attention-based U-structured neural network[47] (right part of Fig. 3) in both stages, so we can extract multi-scale features and ensure that the recovered images of metalens are semantically consistent at various scales, producing high-quality image as expected. Note that the model is optimized differently in the two stages, and 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.[31] Since the design of the metalens is circularly symmetric, it is convenient to rotate these PSFs to obtain approximate PSFs of full fields collection.

$$Image(\lambda)_{meta} = \sum_{\theta} \sum_{\varphi} mask(\theta,\varphi,\lambda) * [psf(\theta,\varphi,\lambda) \otimes Image(\lambda)_{H}] \qquad (1)$$

where $Image(\lambda)_{meta}$ means simulated image corresponding to wavelength λ, $mask(\theta,\varphi,\lambda)$, $psf(\theta,\varphi,\lambda)$ are mask and PSF in theory corresponding to FOV θ, rotation angle φ and wavelength λ respectively, and $Image(\lambda)_{H}$ 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,[32] 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 non-designated wavelengths during our modelling. The detailed processing about PSFs generation can be found in Supplementary Information Section S3. Note that the dataset we collect 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 PSFs set-driven method[31,32] in the second stage to circumvent the following problems. Existing single-wavelength wide FOV metalens with a small front aperture have central bright speckle problem at non-designed wavelengths, which become serious by the increase of incident angle. Unfortunately, so far there are no accurate theoretical model to estimate the central bright speckle. Moreover, the intensity variation and spatial inhomogeneity of PSFs at different angles of incidence and at non-designed wavelengths, making it difficult for the measured PSFs set to restore the real image effect. With such large differences in PSFs intensities, the measured PSFs set ensemble will have a greater loss of precision, resulting in a more tedious and arduous task to measure PSFs 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, imaged by a conventional commercial lens (Sigma Art Zoom lens) or metalens, and finally collects the image pair captured by the CMOS sensor (eg. IMAX 335) and commercial Sony sensor (eg. A7M3, Sony) respectively. More details about this image acquisition processing system can be found in Supplementary Information Section S4. The Supplementary Information Section S5 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 Supplementary Section S7 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 formula:

$$T(I) = T(psf * I) \qquad (2)$$

where T is a particular transformation, I is the imaged object, and PSF is the point spread function 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.[48]

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 1,000 images to validate our approach, 800 as training data and 200 as test data. As shown in Fig. 4(a), the image data of scence (eg. 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

(MPLFLN10xBD, Olympus), an adapter tube lens (1-62922, NAVITAR), and a COMS sensor (A7M3, Sony). Original images captured by the metalens and corresponding recovery results from our neural networks, Unet & KD paradiagm (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 Multi-Scale Retinex with Color Restoration (MSRCR) algorithm are not improved much due to uncorrected background noises. The images recovered by Unet & KD paradiagm 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 (transformer-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) are provided in Section S8 of the supplementary materials. Therefore, our image recovery neural network offer a considerable enhancement in colour 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 × 9.07 × 1.57 mm^3. Fig. 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 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 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 two minutes.

To exhibit the capability of the neural meta-camera, we placed an LCD screen at different working distance 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. Fig. 5(c) shows the results at 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 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 processing error, while the color cast is derived from the difference in spectral response curve of CMOS image sensors between commercial

Sony sensor (A7M3, Sony) and IMX335. Cartoon images from alarm clocks and 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 multi-scale 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 have 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 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 of only meta-camera. Figures 5(e) and 5(f) show the intensity distribution of 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 neural network, respectively. The calculated contrast of 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 actual scene, we captured and recovered the images in two scenarios. One is the imaging of three moniter screens at different working distances; the other is the imaging of multiple objects of various colors arranged at different depths in actual scene. In the first scenario, we obtained recovery images at the working distances of 1.3 cm, 12 cm, and 44.5 cm, as shown in

Figure S17. 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 (PSNR) and structive similarity index measure (SSIM) values (as shown in Table S4) 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 indoor scene. We set up a dual optical path data acquisition system [Figure S18(a)] based on a cube beam splitter to obtain pixel-level aligned datasets. As shown in Figure S18(b), in the recovered image from the neural meta-camera, the letters are clearer, and the dolls at different working distances of 40 cm, 55cm and 85cm 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 actual scene, we compare the performance between imaging of meta-camera and traditional camera on the multi-label image classification task. The data from meta-camera achieves a precision of 96.47%, while the data from traditional camera achieves 96.73%. Experiments demonstrated that imaging of meta-camera did not show significant performance differences in recognition tasks compared to imaging from traditional camera, which hints at the potential of meta-camera for classification and recognition application.

4 Discussion

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. Thanks to the high-precision assembly technology, our neural meta-camera is only 9.07 × 9.07 × 1.57 mm^3 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 wide FOV. Such neural meta-camera is an exemplary case in imaging systems with minimization, 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, etc. 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[49] 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 non-invasive[50] 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.

Acknowledgements

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

Figures and Figure Captions

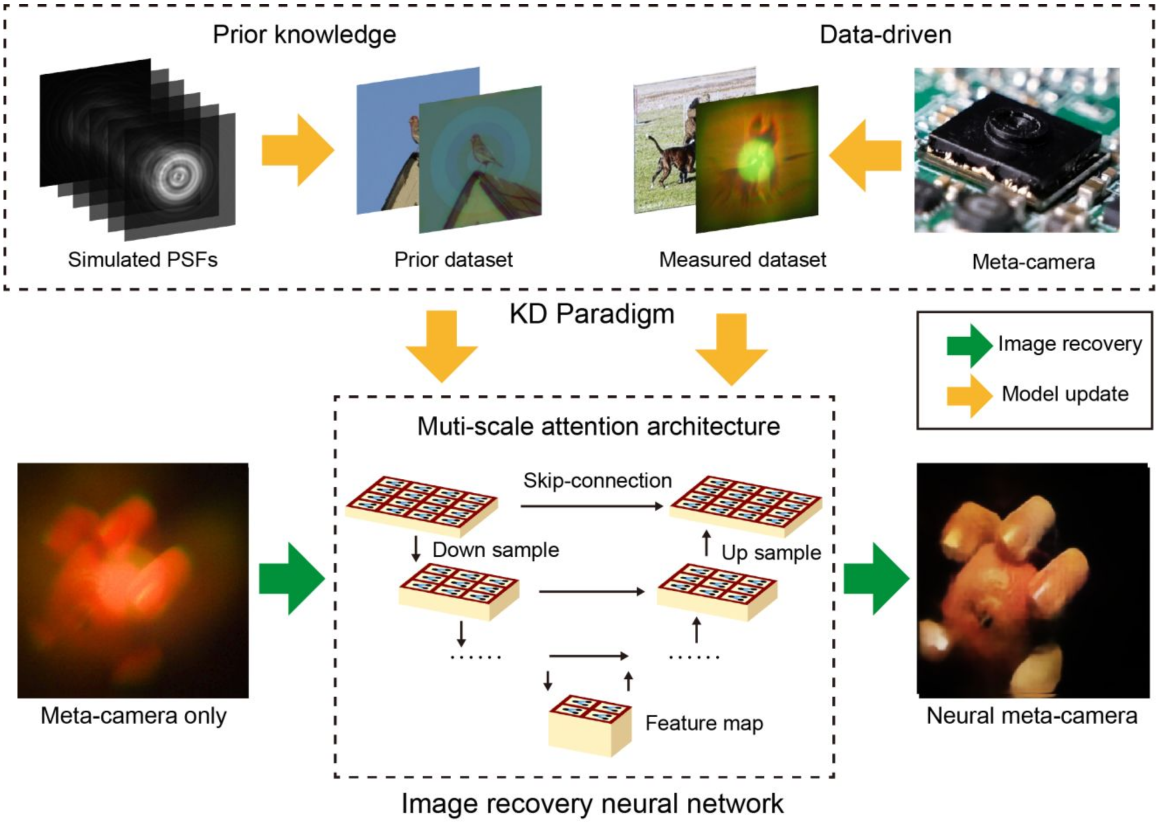

Fig. 1 Neural meta-camera model. The meta-camera consists of the ultra-wide FOV metalens and the transformer-based neural network for full-color imaging. Green arrows show the process of image recovery. The captured image from the meta-camera is reconstructed by the image recovery neural network constructed by the KD paradigm (yellow arrows, prior knowlege and data-driven). The neural network is initialized by the prior dataset from the simulated PSFs of the metalens, and then measured dataset from meta-camera are input to drive the refinement of the initialized neural network. To capture information at muti-scales, 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 non-uniformity.

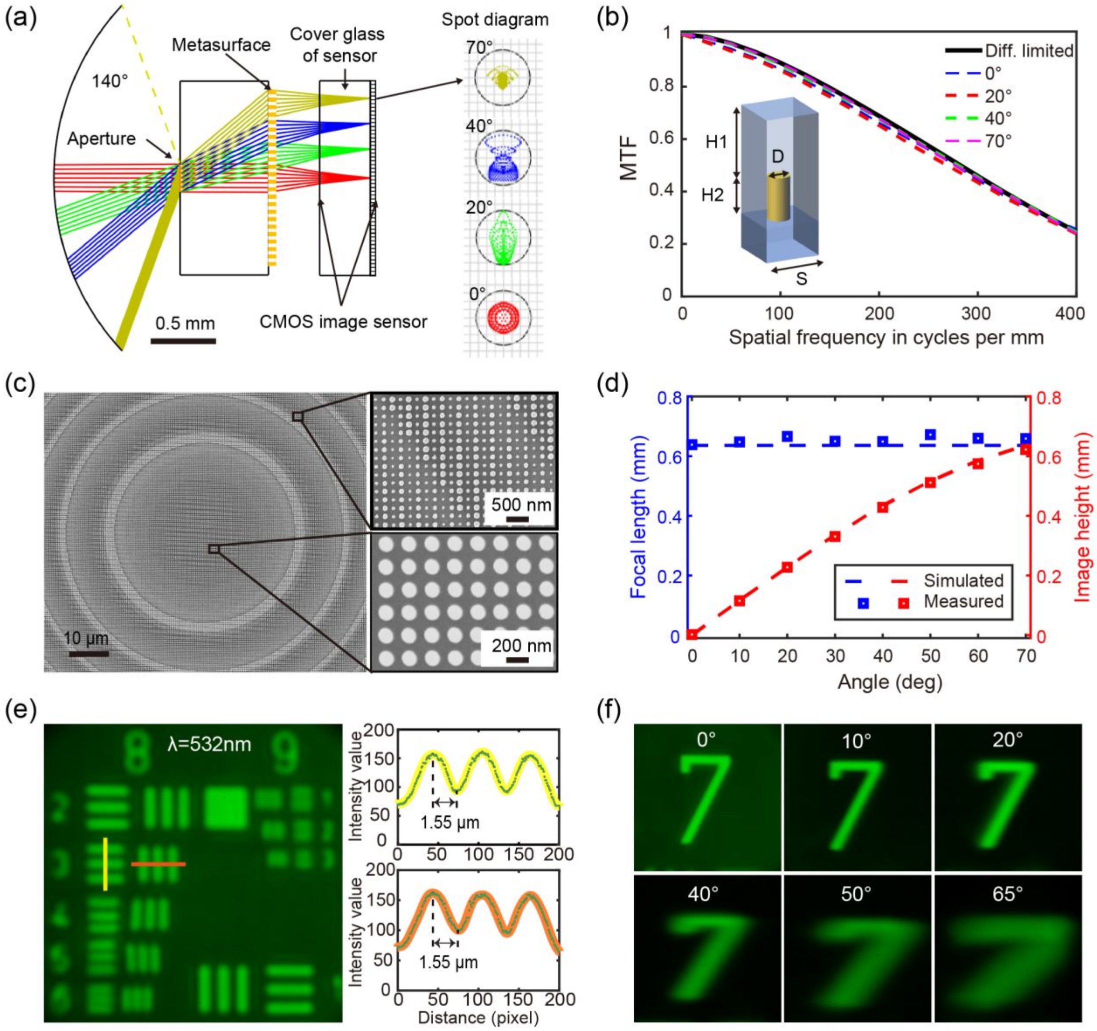

Fig. 2 Ray optics design and characterization of the ultra-wide FOV metalens. (a) Ray tracing simulation results of ultra-wide FOV metalens (left) of 140°. The red/green/blue/yellow rays have four crossing points at the same image plane passing through aperture, substrate, metasurface and cover glass of sensor. Spot diagrams (right) shows the diffuse spots with the incident angles of 0°, 20°, 40°, 70° are inside the Airy circle (black solid). (b) Simulated MTF curves at different FOVs and black solid-line indicates the diffraction limit. Schematic of a meta-atom of the metasurface, consisting of a silicon nanopost with the height (H1) of 265 nm and silicon dioxide protective layer with thinckness (H2) of 1 μm on a silica substrate. The nanoposts with varying diameter (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 the group 8 displayed a line width of 1.55 μm. (f) Image of the number “7” in different incident angles at the wavelength of 532 nm.

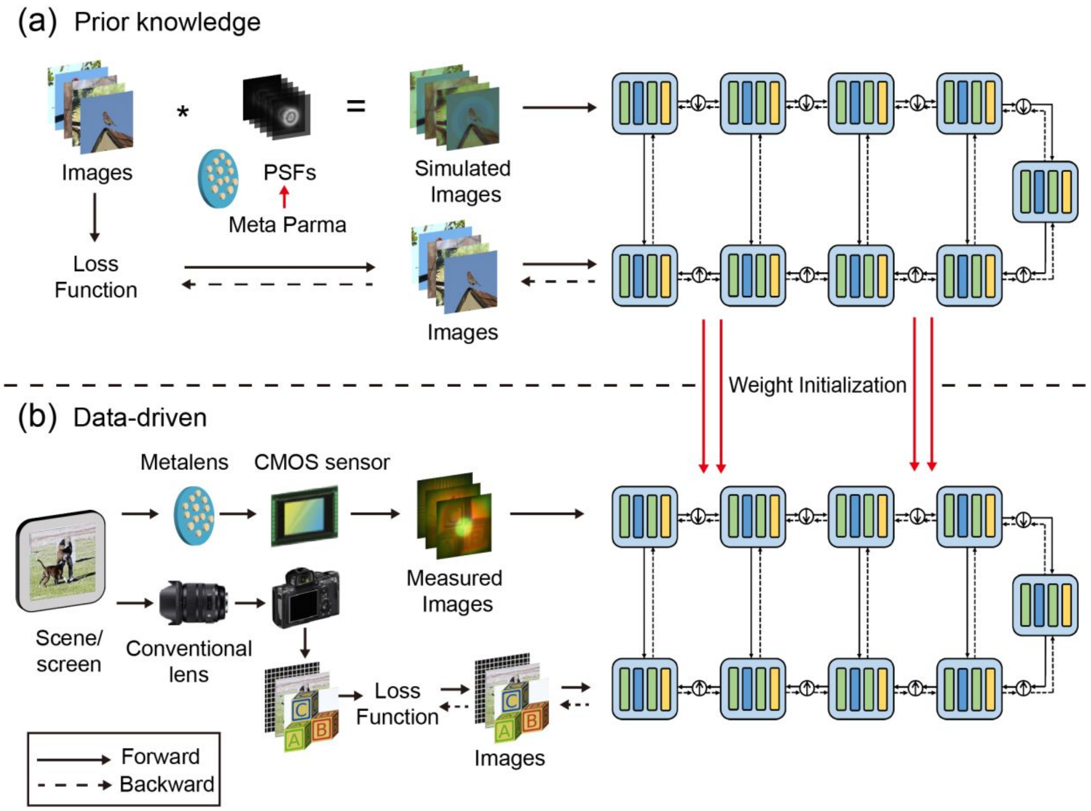

Fig. 3 Proposed KD paradigm for training image recovery neural network. (a) Prior knowledge, i.e. PSFs, obtained from the design parameters of the metalens are applied to the original images to generate the prior dataset. This prior dataset is used to train an initialized neural network. (b) By utilizing the data collection and processing flow we have established, data from corresponding scenarios is collected to drive further fine-tuning of the model, enabling it to cope more intricate image degradation in actual scenarios. Measured dataset in data-driven are images (e.g. LCD screen projection images) captured by metalens and conventional commercial lens (Sigma Art Zoom lens) respectively. As shown by the black dotted line, the neural network is updated through back-propagation with 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.

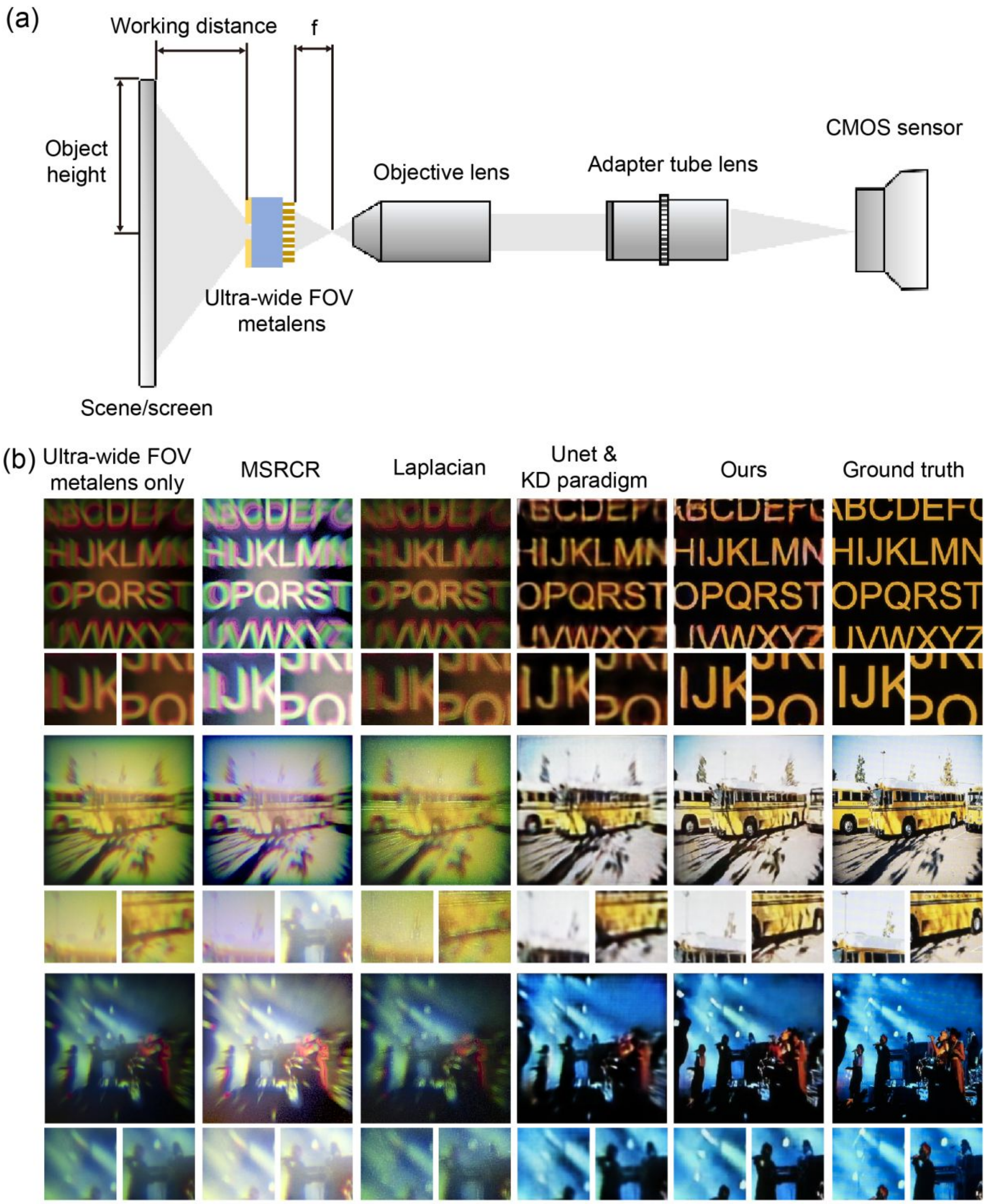

Fig. 4 Image recovery results of our neural network for images of naked ultra-wide FOV metalens are compared with results from Unet & KD paradigm and other traditional method. (a) Schematic illustrations of data acquisition system for naked ultra-wide FOV metalens. The object projected by a 5.5-inch LCD screen is collected by the naked ultra-wide FOV metalens with a working distance of 2 cm and redirected to a micromagnification system with an objective lens (Olympus, MPLFLN10xBD), adapter

tube lens (1-62922, NAVITAR) and a CMOS sensor (Sony, A7M3). (b) Compared to Unet & KD paradiagm 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, concerts under blue lights. Detail insets are illustrated below each row. Compared to ground truth capture (the right most column) using 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, oranges) are shown in Fig. S12-14.

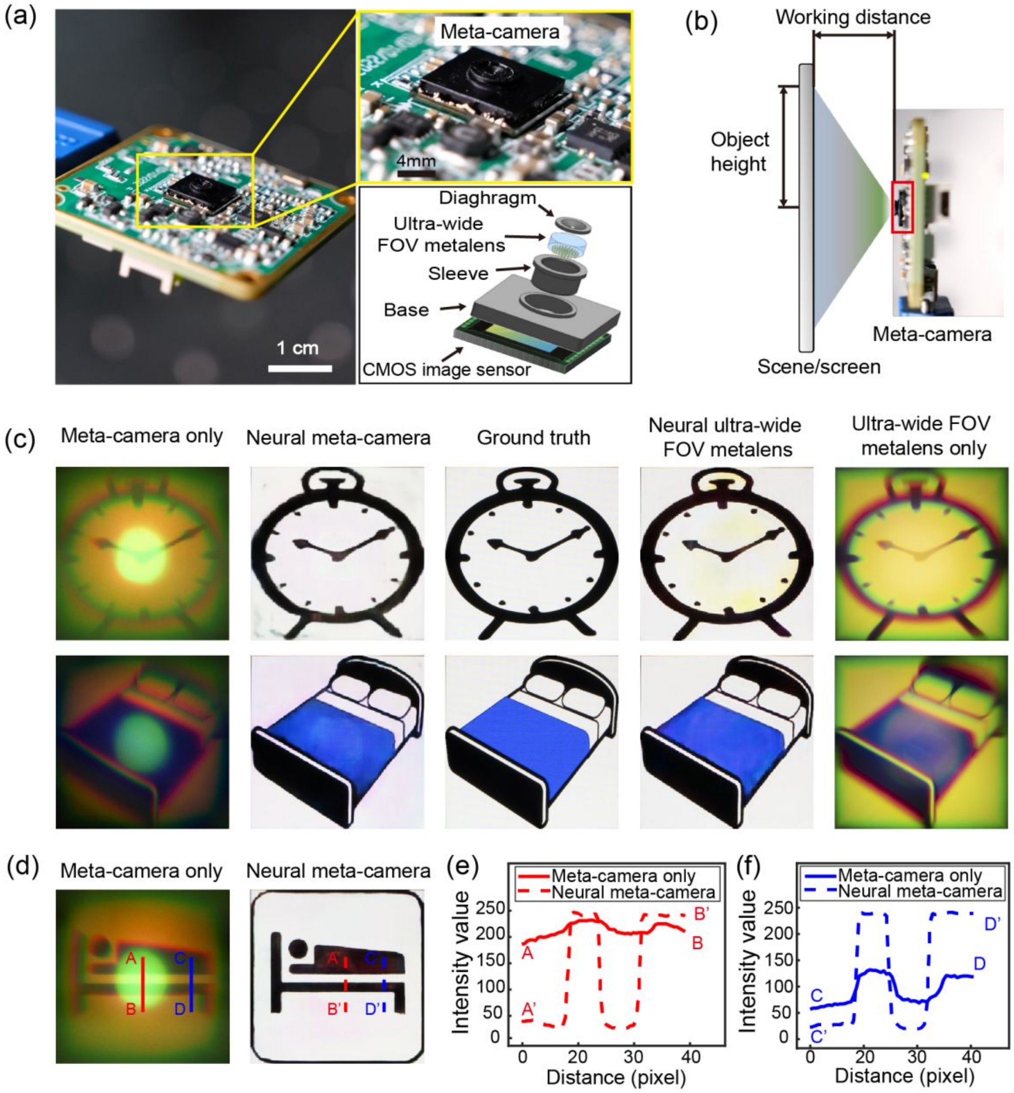

Fig. 5 Neural meta-camera for imaging. (a) Photograph of the meta-camera system (left) by integrating the miniature meta-camera (top-right) with a CMOS image sensor, and the schematic illustrations of its structural mechanism (bottom-right) includes an aperture, sleeve and base for shading and waterproofing. (b) Schematic diagram of meta-camera test. The ground truth images are projected on the LCD screen and captured directly by meta-camera. (c) Comparison recovery results from images captured by ultra-wide FOV metalens only and meta-camera at the working distance of 2cm. Cartoon images from alarm clocks and 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, concert) are shown in Fig. S15-16. (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.

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 transmitarrays," *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., "N. 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.* **2**, 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. Tseng et al., “Neural nano-optics for high-quality thin lens imaging,” *Nat. Commun.* **12***,* 6493 (2021).

32 Q. Fan et al., “Trilobite-inspired neural nanophotonic light-field camera with extreme depth-of-field,” *Nat. Commun.* **13**, 2130 (2022).

33 Y. Peng et al., “The diffractive achromat full spectrum computational imaging with diffractive optics,” *ACM Trans. on Graph.* **35**, 31 (2016).

34 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. on Graph.* **37**, 114 (2018).

35 Y. Peng et al., "Learned large field-of-view imaging with thin-plate optics," *ACM Trans. on Graph.* **38**, 1-14 (2019).

36 L.Yann, B. Yoshua, and H. Geoffrey, "Deep learning," *Nature* **521**, 436-444 (2015).

37 J. Jumper et al., "Highly accurate protein structure prediction with AlphaFold," *Nature* **596**, 583-589 (2021).

38 T. Brown et al., "Language models are few-shot learners," *Adv. Neural Inf. Process.- Syst.* **33**, 1877-1901 (2020).

39 S. Pinilla et al., "Miniature color camera via flat hybrid meta-optics," *Sci. Adv*. **9**, eadg7297 (2023).

40 S. Molesky et al., "Inverse design in nanophotonics," *Nat. Photonics* **12**, 659-670 (2018).

41 W. Ma et al., "Deep learning for the design of photonic structures," *Nat. Photonics* **15**, 77-90 (2021).

42 F. Wang et al., "Phase imaging with an untrained neural network," *Light Sci. Appl.* **9**, 77 (2020).

43 V. Liu and S. Fan, "S4 : A free electromagnetic solver for layered periodic structures," *Comput. Phys. Commun.* **183**, 2233–2244 (2012).

44 A. Vaswani et al., "Attention is all you need," *In Proceedings of the 31st Annual Conference on Neural Information Processing Systems (NIPS)* pp. 6000-6010 (2017).

45 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(MICCAI)* pp. 234-241 (2015).

46 C. Tan et al., "A survey on deep transfer learning," *In International conference on artificial neural networks(ICANN)* pp. 270-279 (2018).

47 Z. Wang et al., "Uformer: A general u-shaped transformer for image restoration," *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR)* pp. 17683-17693 (2022).

48 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 Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR)* pp. 5647-5656 (2022).

49 J. Xiong et al., "Dynamic brain spectrum acquired by a real-time ultraspectral imaging chip with reconfigurable metasurfaces," *Optic*a **9**, 461-468 (2022).

50 L. Zhu et al., "Large field-of-view non-invasive imaging through scattering layers using fluctuating random illumination," *Nat. Commun.* **13**, 1447 (2022).

51 E. E. Fenimore, "Coded aperture imaging: the modulation transfer function for uniformly redundant arrays," *Appl. Opt.* **19**, 2465-2471 (1980).

52 S. R. Gottesman and E. E. Fenimore, “New family of binary arrays for coded aperture imaging,” *Appl. Opt.* **28**, 4344-4352 (1989).

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).