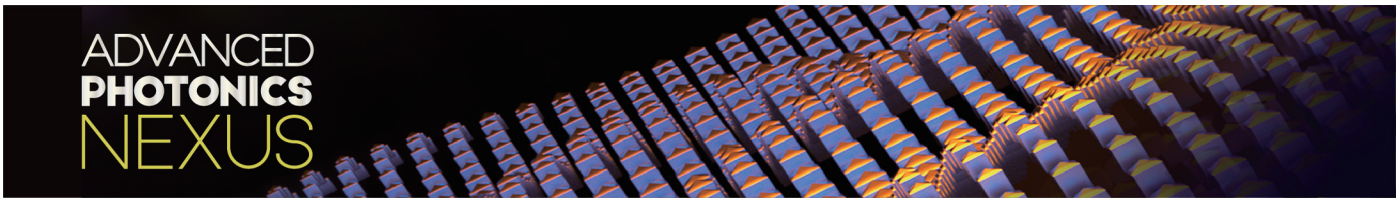
Research Article



Neural nanophotonic object detector with an ultrawide field-of-view

Ji Chen,[a,b,*,†] Muyang Li,[a,†] Zhongyi Yuan,[a,†] Yue Wu,[a,†] Zi-Wen Zhou,[a] Bingcheng Zhu,[a,b] Yin Wang,[b] Jitao Ji,[c] Chunyu Huang,[c] Kai Qiu,[c] Shining Zhu,[c] Tao Li,[c,*] and Zaichen Zhang[a,b,*]

[a]Southeast University, School of Information Science and Engineering, Frontiers Science Center for Mobile Information Communication and Security, Quantum Information Research Center, National Mobile Communications Research Laboratory, Nanjing, China
[b]Purple Mountain Laboratories, Nanjing, China
[c]Nanjing University, College of Engineering and Applied Science, School of Physics, National Laboratory of Solid State Microstructures, Nanjing, China

Abstract. Intelligent object detection, which extracts crucial information such as object categories and locations, plays a vital role in emerging technologies, including autonomous driving, the Internet of Things, and mobile sensing. Achieving high detection performance over a wide angular range typically demands complex and bulky optical systems, significantly limiting the detectors' applicability. Here, we present an ultracompact neural nanophotonic object detector based on a metalens array, achieving high-accuracy object recognition (>96%) and high-precision localization (<0.1 deg) across an ultrawide field-of-view (>135 deg) under static conditions. Benefiting from its high compactness and planar architecture, we integrated the device into a miniature unmanned aerial vehicle, enabling real-time intelligent detection across diverse real-world environments with exceptional flexibility and mobility. Our work fully leverages the ultrathin and lightweight advantages of the metalens and demonstrates a full-stack design-to-system pipeline, spanning nanostructure fabrication, neural algorithm development, and system-level integration, offering a scalable framework for next-generation intelligent photonic sensing systems.

Keywords: metalens array; wide-angle imaging; neural network; integrated imaging device.

Received Nov. 30, 2025; revised manuscript received Apr. 16, 2026; accepted for publication May 6, 2026; published online May 27, 2026.

[DOI: 10.1117/1.APN.5.4.046013]

1 Introduction

Intelligent object detection refers to automatic identification of objects of interest within a given scene, enabling subsequent information processing or interaction with the identified targets. It has broad applications across various fields, including autonomous driving, security and surveillance, industrial automation, and smart cities.[1–5] Among the available approaches, imaging-based methods powered by computer vision have emerged as a dominant technique due to their numerous advantages, such as high resolution, fast processing speeds, and rich information transmission.[6–10] In scenarios requiring high portability or operations in confined spaces, ultracompact imaging devices are very crucial. However, size constraints often result in trade-offs in imaging performance, particularly in terms of image quality and field-of-view (FOV), which would significantly affect the accuracy and effective range of subsequent object detection processes.

Metalenses are artificial lenses composed of two-dimensional (2D) subwavelength units.[11–16] Thanks to their ultralightweight, ultrathin, and flexible design features, metalenses have been successfully employed to create highly integrated imaging devices with diverse functionalities.[17–22] Certain high-efficiency dielectric metalenses can deliver performance comparable to that of commercial objective lenses.[23–27] However, achieving a large FOV remains a significant challenge for metalenses. Although several wide-FOV metalens designs have been proposed, including two-layer metalenses, quadratic phase metalenses, and computational thin-plate lenses,[28–36] these methods often encounter fabrication or integration difficulties or compromise key imaging

*Address all correspondence to Ji Chen, jichen@seu.edu.cn; Tao Li, taoli@nju.edu.cn; Zaichen Zhang, zczhang@seu.edu.cn
†These authors contributed equally to this work.

performance metrics such as efficiency or resolution. Developing a single-layer metalens with capabilities of both ultrawide FOV and high imaging quality is crucial for advancing highly integrated intelligent object detectors, allowing them to detect a wider range of targets effectively in a compact form.

In this work, we propose a neural nanophotonic object detector (NNOD) based on a single layer of a metalens array and a comprehensive neural network, achieving an ultrawide FOV of 135 deg [Fig. 1(a)]. Three metalenses of the metalens array are meticulously designed to provide clear imaging for different angular ranges [Fig. 1(b)]. The sub-images produced by these metalenses collectively form a complete wide-angle image through processes of distortion correction and stitching. A comprehensive neural network is then employed to further enhance the captured image's quality and intelligently identify target objects in the image [Fig. 1(c)]. These advancements enable the NNOD to demonstrate high object detection precision (>96%), large detection distance (>10 m), and precise angular localization under static conditions (<0.1 deg). In addition, we presented an integration solution of the NNOD with a miniature unmanned aerial vehicle (MAV), featuring wireless transmission of captured images of diverse real-world environments, and high-performance real-time intelligent detection of the images' contents. Our study presents a systematic framework for advancing revolutionary intelligent detection systems, offering significant potential for a wide range of future applications.

2 Results

2.1 Design of the Wide-Angle Metalens Array

The core component of the NNOD is a metalens array for wide-angle imaging, comprising three metalenses designed for different viewing angles (one center-FOV metalens and two side-FOV metalenses). However, directly designing a metalens with a broad viewing angle is inherently challenging as it requires the unit cells to adjust phase responses to varying angles of incidence (see Sec. S1 in the Supplementary Material for details). To overcome this challenge, we first employed a multi-configuration ray tracing optimization method to derive the phase profile of the center-FOV metalens [Fig. 2(a)]. The basic concept of this approach is to optimize the phase profile by defining specific virtual apertures that direct incident light from various angles to illuminate distinct regions of the metalens (see Sec. 4 for optimization details). It should be noted that the apertures used in optimization are virtual, rather than physical. When the virtual apertures are removed from the ray tracing model, the actual light modulation of the designed single-layer center-FOV metalens is revealed [Fig. 2(b)]. Light passing through the virtual aperture regions is well focused by the metalens, with the focal spot size comparable to that of the Airy disk, as shown in Fig. 2(a). Light outside the virtual aperture regions is evenly distributed across the complementary metal oxide semiconductor (CMOS) sensor, which would be effectively mitigated in subsequent neural network processing. Although adding a physical aperture mask at the virtual aperture positions could further block ambient light and improve image quality, we opted for an aperture-stop-free design to avoid additional fabrication complexity and alignment challenges. After a comprehensive evaluation, we determined the center-FOV metalens features a diameter of 1 mm, a focal length of 1.6 mm, a working wavelength of 550 nm, and an FOV ranging from −22.5 to 22.5 deg, respectively. See Secs. S2 and S3 in the Supplementary Material for detailed analysis of the parameters. The simulated focusing performance of metalenses with different FOVs is shown in Fig. S2 in the Supplementary Material.

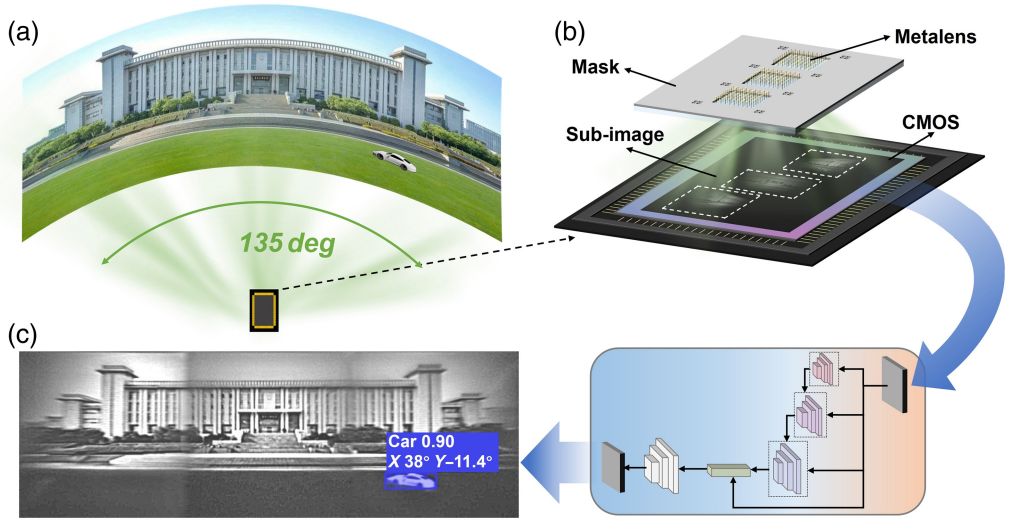


Fig. 1 Working principle of the neural nanophotonic object detector (NNOD). (a) The NNOD is capable of imaging a wide scene with an FOV exceeding 135 deg. (b) The key hardware components of NNOD are a metalens array and a complementary metal–oxide–semiconductor (CMOS) sensor. The three metalenses of the metalens array produce clear images of objects from different viewing angles, creating three sub-images. The surrounding mask blocks direct ambient light to enhance image quality. (c) Three sub-images are then fed into a comprehensive neural network to get a high-quality wide-angle imaging result. The target objects within the wide-angle image are accurately identified and located.

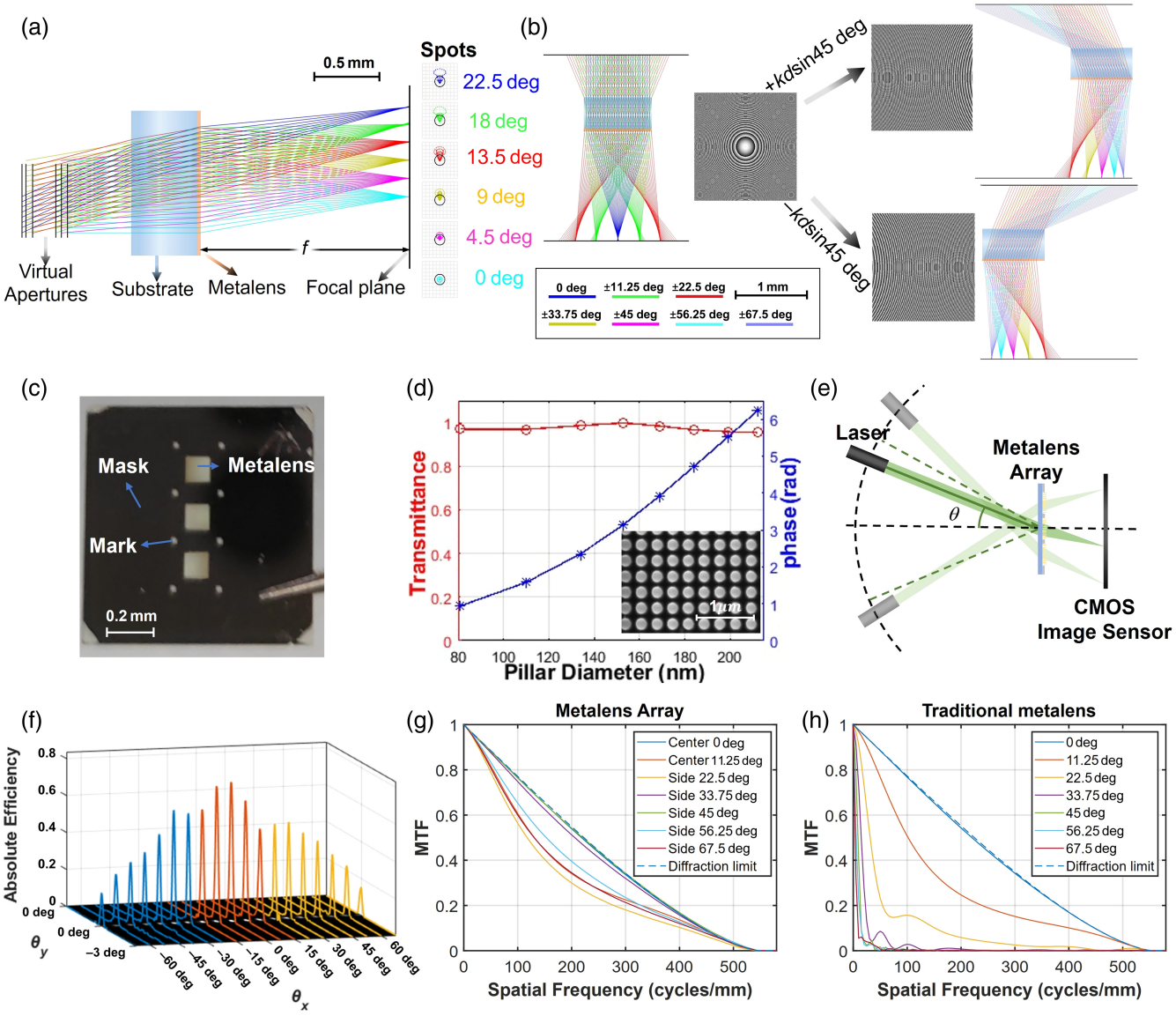


Fig. 2 Design and characterization of the wide-angle metalens array. (a) Ray tracing simulation of the center-FOV metalens with virtual apertures. The focal spots of light rays with incident angles of 0, 4.5, 9, 13.5, 18, and 22.5 deg are shown on the right side. The black circles indicate the size of the Airy disks corresponding to different incident angles. (b) Ray tracing simulation of the center-FOV metalens and the two side-FOV metalenses without virtual apertures. The phase profiles of the two side-FOV metalenses are obtained by adding the tilt phase $\pm kd$ sin 45 deg to the center-FOV metalens phase profile, respectively. (c) Photograph of the metalens array sample. The three large square regions represent the metalens structure areas, whereas the small regions around the metalenses are markers for alignment in fabrication. The surrounding black areas correspond to the mask. (d) Phase (blue stars) and transmittance (red circles) of meta-atoms with eight different pillar diameters (81, 110, 134, 153, 169, 184, 199, and 212 nm), simulated at the wavelength of 550 nm. The inset is the top-view scanning electron microscope (SEM) image of part of the nano-structures. (e) Optical setup for characterization of metalens array focusing performance. (f) Focal spots and their absolute efficiency at different incident angles. The modulation transfer functions (MTFs) of (g) the wide-angle metalens array and (h) a traditional hyperbolic phase metalens.

The phase profiles of the two side-FOV metalenses require no further optimization; instead, they can be obtained by adding a tilt phase $\pm kd$ sin θ_0 to the center-FOV metalens phase, where k is the wave number, d is the metalens diameter, and θ_0 is the designed tilt angle. Introducing the tilt phase effectively compensates for the additional phase shifts caused by oblique incidence, thereby enabling a precise adjustment of the metalens imaging angular range. Here, θ_0 is set as 45 deg to make the two side-FOV be −67.5 to −22.5 deg and 22.5 to 67.5 deg, to symmetrically offset the center-FOV. The experimental imaging

results of metalens arrays designed for different FOVs are presented in Fig. S3 in the Supplementary Material, which further validates the choice of a 135 deg total FOV. Ray tracing analysis of the two side-FOV shows that major rays are focused to high-quality spots such as the center-FOV metalens [Fig. 2(b)], ensuring the wide-angle imaging capability of the entire metalens array.

2.2 Fabrication and Optical Characterizations of the Wide-Angle Metalens Array

The metalens array was fabricated on a silica (SiO_2) substrate, arranged longitudinally with their imaging FOV oriented laterally to minimize the cross-talk among the sub-images. To further reduce noise from direct light transmission outside the metalenses structure regions, a mask was fabricated around the metalenses, as shown in Fig. 2(c). The metalenses' structures were fabricated in silicon nitride (Si_3N_4) nano-posts with a period of 275 nm, a height of 1000 nm, and varying radius of round cross-sections. This parameter configuration effectively suppresses higher-order diffraction, as detailed in Sec. S4 in the Supplementary Material. Through finite-difference time-domain (FDTD) simulation, eight types of nano-posts with varying cross-sectional radii were identified, which cover the 2π phase range and all exhibit over 90% transmittance [Fig. 2(d)]. The inset displays the top-view of a scanning electron microscope (SEM) image of the metalens structures. See Sec. S5 in the Supplementary Material and Sec. 4 for details of the metalens structures simulation and fabrication, respectively. Although the FDTD simulations of the nano-posts were performed only at a wavelength of 550 nm and under normal incidence (0 deg), their transmittance and phase delay remain stable over a broad wavelength range and under oblique incidences, as detailed in Sec. S6 in the Supplementary Material. Figure S9 in the Supplementary Material further demonstrates the angular stability of the meta-atoms, showing their transmission and phase delay under different incident angles.

Upon completion of the metalens array fabrication, optical testing was conducted to verify its wide-angle focusing performance, the schematic diagram of the experimental setup of which is shown in Fig. 2(e). A 550-nm light filtered from a supercontinuum laser (NKT, SuperK COMPACT, Copenhagen, Denmark) is accurately controlled by an optical stage. The optical stage rotates along an arc with a radius of 50 cm, centered on the metalens array sample, and the emitted laser continuously faces the sample. When the laser rotates to different angles, the optical stage is adjusted accordingly to allow different metalenses to receive the light. The two side-FOV metalenses receive light at angles from −67.5 to −22.5 deg and from 22.5 to 67.5 deg, respectively, whereas the center-FOV metalens receives light at angles from −22.5 to 22.5 deg. After the laser beam is focused by the metalens array, the focal spots are captured by the CMOS image sensor (Imaging source, DMM 27UJ003-ML, Bremen, Germany) for further analysis. The focal spots and absolute focusing efficiency of light corresponding to different incident angles are shown in Fig. 2(f). The narrow full width at half-maximums (FWHMs) of the curves indicate that the focal spots maintain good focusing quality. The modulation transfer functions (MTFs) of the metalens array were also obtained through experimental measurements [Fig. 2(g)]. Compared with the MTFs of a traditional hyperbolic phase metalens[35] [Fig. 2(h)], the MTFs of our metalens array for different incident angles closely approach the diffraction limit curve, demonstrating significantly improved wide-angle imaging capability.

2.3 Wide-Angle Imaging of the Metalens Array

We initially conducted imaging experiments on the optical platform to determine the parameters for processes of distortion correction and sub-image stitching, and also to acquire the imaging datasets for subsequent neural network training, the experimental setup of which is shown in Fig. 3(a). The wide-angle images were projected onto a display composed of three organic light-emitting diode (OLED, CFORCE, CF015Next) screens. The screens were mounted on adjustable stands, allowing for continuous control of the entire viewing angle by adjusting the relative positions between each pair of screens. The imaging device composed of the metalens array and the CMOS was mounted on a metal support rod to align with the center height of the screen and was connected to a desktop computer via a high-speed transmission line, enabling real-time transmission of the captured images. Imaging distortion is an inherent challenge in wide-angle imaging, and effectively mitigating it is essential for improving both imaging quality and the accuracy of intelligent object detection and localization. Through the imaging distortion analysis of a 2D dot matrix and a chessboard pattern, a distortion correction method was successfully developed (see Sec. 4 and Sec. S7 in the Supplementary Material for details). As imaging distortion is primarily dependent on the imaging device, the distortion mapping is fixed once the metalens array is fabricated. Consequently, the distortion correction method can be applied to any subsequent wide-angle images of the same metalens array.

We first conducted wide-angle imaging of a pure letter pattern of "Southeast University" with a viewing angle of 150 deg. Due to its simplicity and high contrast, the imaging viewing angle could exceed the designed FOV of 135 deg. The sub-images captured directly by the CMOS exhibit obvious imaging distortion with distinct regions of clear and blurred imaging [Fig. 3(b)]. After distortion correction, the clear imaging regions of every sub-image were extracted and stitched together to produce the complete wide-angle image, as shown in Fig. 3(c). Notably, the simple shape and content of the letter pattern allow high-quality imaging to be achieved without further processing. For comparison, a traditional hyperbolic phase metalens was also employed to capture the image, the result of which is shown in Fig. 3(d). It is evident that our wide-angle metalens array expands the FOV by nearly three times. In addition, we imaged a complex scene [the original image shown in Fig. 1(a)], which imposes much higher demands on the imaging capability. Thus, we set the imaging angle to the designed 135 deg, and following the distortion correction and stitching processes, we can obtain the wide-angle imaging result of this complex scene, as shown in Fig. 3(e).

2.4 Intelligent Image Processing and Object Detection

Image quality is essential in object detection performance,[8,37,38] with higher-quality images typically yielding greater detection accuracy. Although our metalens array has produced clear, wide-angle images, there remains room for further improvement. Factors such as light transmission outside the virtual apertures, environmental stray light, and intensity attenuation at large angles can all contribute to the degradation of image quality. To address these issues, we proposed a comprehensive neural network, incorporating an adaptive image quality enhancement module (module 1) and an intelligent image content recognition module (module 2). The concise schematic diagram of the neural network structure is shown in Fig. 3(f).

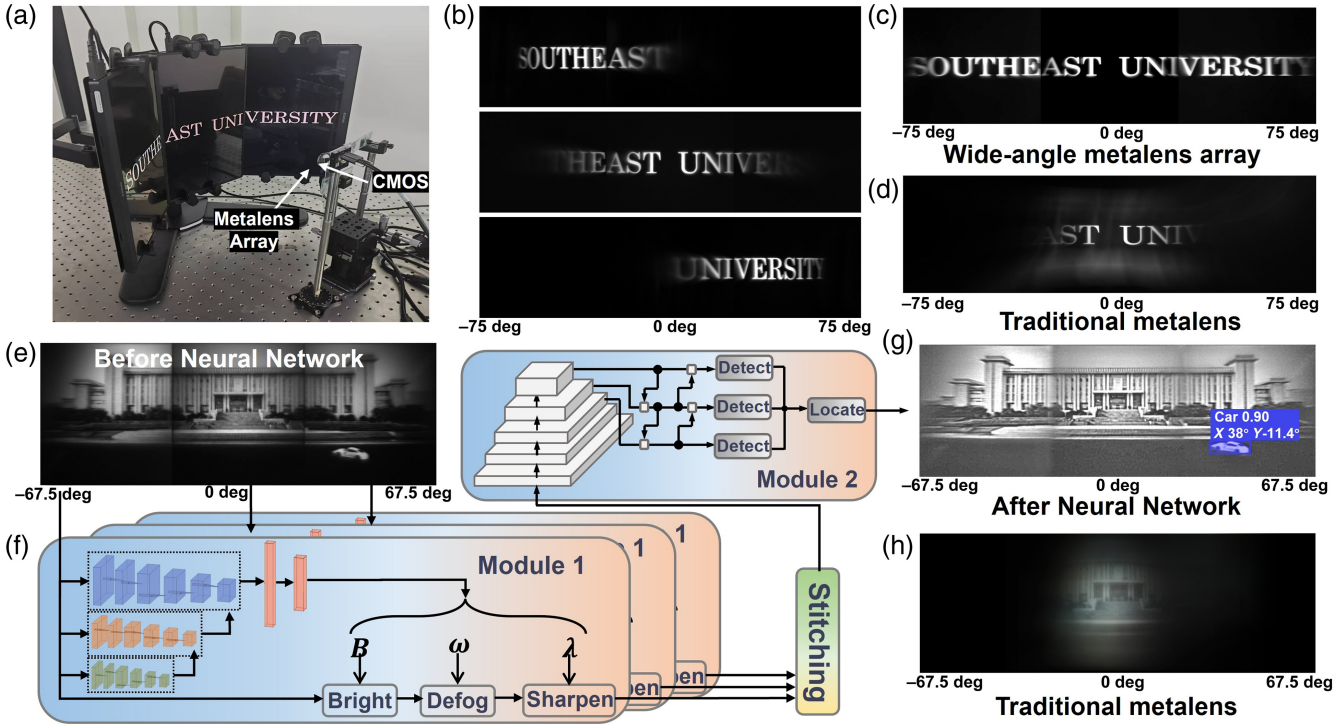


Fig. 3 Wide-angle imaging and neural network processing by the NNOD. (a) The setup for wide-angle imaging and neural network data collection on the optical platform. The images are projected onto a wide-angle display consisting of three OLED screens. The positions of the CMOS and metalens are adjusted to set the distance between them equal to the metalens focal length. (b) Sub-images directly captured by the three metalenses, each of which has its own clear imaging parts. (c) Wide-angle image of the letter pattern obtained by stitching the clear image part of each sub-image, with a total FOV of 150 deg. (d) Imaging result of the same letter pattern by a traditional hyperbolic phase metalens, the FOV of which is limited. (e) Wide-angle imaging result of a complex image, with an FOV of 135 deg. (f) Neural network processing. The three sub-images are initially processed by module 1 for image quality enhancement. Subsequently, these images are stitched as a wide-angle image, which is then fed into module 2 for intelligent object recognition and localization. (g) The final wide-angle image and intelligent detection results after the neural network processing. The image highlights the recognition confidence of the target object and its x and y angle localization coordinates. (h) Imaging result of the same complex scene by a traditional metalens for imaging FOV and quality comparison.

Module 1 includes a multiscale convolutional neural network (MSCNN) to extract hyperparameters (brightness, contrast, dynamic range, noise, etc.) of captured images, which would be sent to the subsequent multiple filters for image processing (the detailed structure and working principle of module 1 are shown in Sec. S8 in the Supplementary Material). The three sub-images were processed individually by module 1 and then stitched together to form a complete high-quality wide-angle image. Before being fed into module 2, a stitching artifact reduction operation is applied to further improve the image continuity (see Sec. S9 in the Supplementary Material for details). Module 2 is an improved network based on the YOLO network,[39–42] incorporating a high-precision angular localization function. This module divides the input image into a grid and simultaneously predicts for each grid cell the bounding box coordinates, a confidence score indicating the likelihood of containing an object, and the associated class probabilities. The final prediction is made by selecting the bounding box with the highest confidence score for each detected object, and the angular localization is computed based on the predicted coordinates of the object's bounding box (the detailed structure and working principle of module 2 is shown in Sec. S10 in the Supplementary Material).

Through the neural network processing, not only a much higher quality wide-angle image is obtained, but also the objects in it can be intelligently detected, as shown in Fig. 3(g). Compared with the traditional metalens imaging result [Fig. 3(h)], the detection information is greatly enhanced. To demonstrate the superior performance of the neural network, more comparison results of wide-angle images before and after neural network processing are presented in Sec. S11 in the Supplementary Material. Before conducting detection in real scenarios, we trained NNOD to possess the capability of recognizing various categories of objects. Section S12 in the Supplementary Material shows the detailed training process. To test the training effect, wide-angle images containing various objects (common categories including cars, persons, trucks, buses, UAVs, and bicycles) were projected on the wide-angle display for detection. The intelligent detection results demonstrate that

the NNOD achieves a high recognition accuracy for all targets, on average exceeding 96% (see Sec. S13 in the Supplementary Material for details). Video S1 shows the wide-angle detection results of a home-made video, which contains various dynamic common types of objects (Video S1, MP4, 4.71 MB [URL: https://doi.org/10.1117/1.APN.5.4.046013.s1]).

2.5 High-Accuracy Angular Positioning Performance of the NNOD

In addition to intelligent recognition of object categories within the scene, the NNOD is also capable of accurately determining the position of each target, which is represented in terms of angular coordinates. The positioning function is achieved by extracting the object's angular coordinates from the bounding box identified by the neural network module 2. If the bounding box is parameterized by (x, y, w, h), where (x, y) denote the center location and w, h denote the width and height, respectively, the 2D angular position of the target can be derived by

$$\text{angle}_x = \begin{cases} \tan^{-1}\left(\frac{\tan(a)}{p}(x-p)\right) - 2a, 0 < x < 2p \\ \tan^{-1}\left(\frac{\tan(a)}{p}(x-3p)\right), 2p < x < 4p \\ \tan^{-1}\left(\frac{\tan(a)}{p}(x-5p)\right) + 2a, 4p < x < 6p \end{cases},$$

$$\text{angle}_y = \tan^{-1}\left(\frac{\tan(a)}{p}(p-y)\right), \quad (1)$$

where p is half of the image horizontal size and a is one-sixth of the FOV. To ensure the positioning accuracy, it is necessary to first pre-calibrate the positions of the objects in wide-angle images. Although distortion recovery algorithms have been employed in the wide-angle imaging result, there is still a slight deviation in angle calculation. To mitigate this deviation, we implemented a compensation algorithm. The target object car is placed at 20 different positions on the original image, and the wide-angle imaging results are captured and processed to remove distortion. Then, both the original images and the processed imaging results are subjected to the target recognition algorithm, producing 20 pairs of coordinates. The coordinates identified in the original images can be considered as the ground truth, whereas those obtained from the processed imaging results will exhibit certain deviations. These deviation values are then interpolated to create an error compensation curve across the full viewing angle range. Through this curve, the compensation angle for the coordinate position at any given angle can be obtained, by adding which, the accurate angular localization results would be achieved.

After finishing the angular position calibration, we first tested the positioning capability by randomly putting the object car at four representative angular positions $(-55.6, -12.3$ deg), $(-9.9, -8.4$ deg), $(9.1, -8.6$ deg), and $(38, -11.4$ deg), respectively [Fig. 4(a)]. The two coordinates represent the angular positions in the x- and y-directions, respectively. To ensure the reliability of the NNOD, we manually inspected the imaging, recognition, and processing results to confirm that the detected bounding boxes accurately enclosed the target objects before conducting the angular localization experiments. After the imaging, recognition, and processing of NNOD, and adding the aforementioned angular compensation, the positioning results of the four cars are completely consistent with the set angle coordinates, as shown in Fig. 4(b). Second, to further demonstrate the high accuracy of our positioning method, the object car was put at four other positions; the angular positions at x

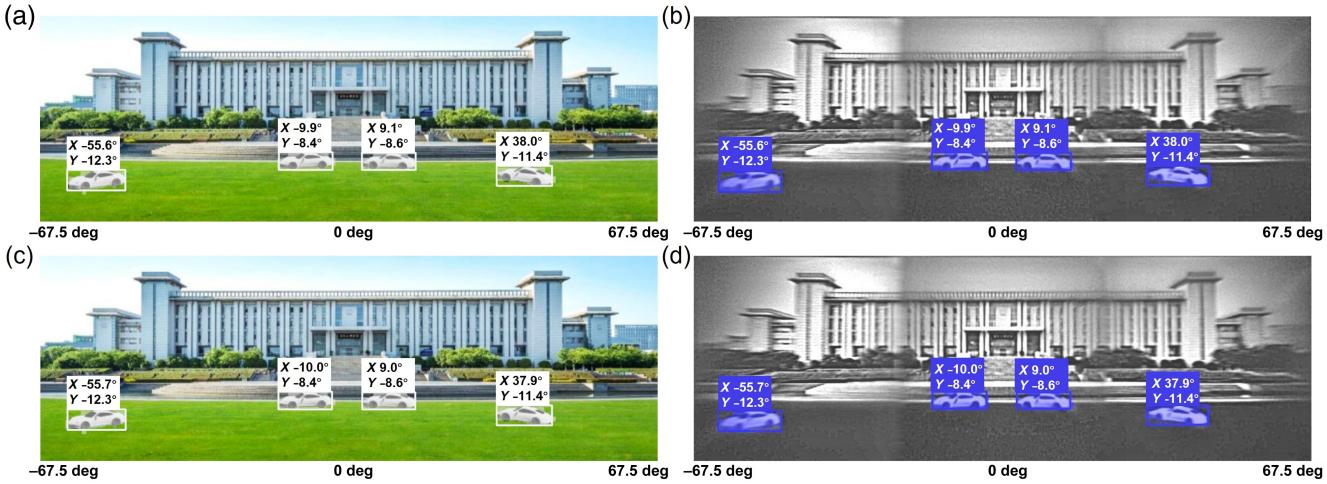


Fig. 4 High-precision angular positioning performance of the NNOD. (a) Original wide-angle image where the target car is placed at four different angular positions (–55.6, –12.3 deg), (–9.9, –8.4 deg), (9.1, –8.6 deg), and (38, –11.4 deg), respectively, including both large and small angular positions. (b) The NNOD detection results of panel (a) after angular compensation. The detected angular coordinates of the cars are completely consistent with those set in the original image. (c) Original wide-angle image where the target car is placed at four other different angular positions (–55.7, –12.3 deg), (–10.0, –8.4 deg), (9.0, –8.6 deg), and (37.9, –11.4 deg), respectively, differing by 0.1 deg in x-direction from those in panel (a). (d) The NNOD detection results of panel (c). The results show that even though the cars in the two original images differ by only 0.1 deg, they can still be accurately detected by the NNOD.

are 0.1 deg different from the previously set positions, which are $(-55.7, -12.3$ deg$)$, $(-10.0, -8.4$ deg$)$, $(9.0, -8.6$ deg$)$, and $(37.9, -11.4$ deg$)$, respectively [Fig. 4(c)]. The detection results in Fig. 4(d) demonstrate that, under static conditions, NNOD can accurately distinguish angular position differences as small as 0.1 deg, a level of precision imperceptible to the human eye, highlighting its outstanding localization capability. Further details on the relationship between localization accuracy and viewing angle are presented in Sec. S14 in the Supplementary Material.

2.6 High-Mobility Detection Based on an NNOD-Integrated MAV System

We then integrated the NNOD into an MAV (Fancinnov, FanciSwarm, Zhejiang, China), enabling flexible wide-angle detection across various real-world scenarios, as shown in Fig. 5(a). The NNOD, along with other components such as the power supply module and signal transmission module, is mounted onto the MAV using a 3D-printed bracket (see Sec. S15 in the Supplementary Material for details of integration). The NNOD

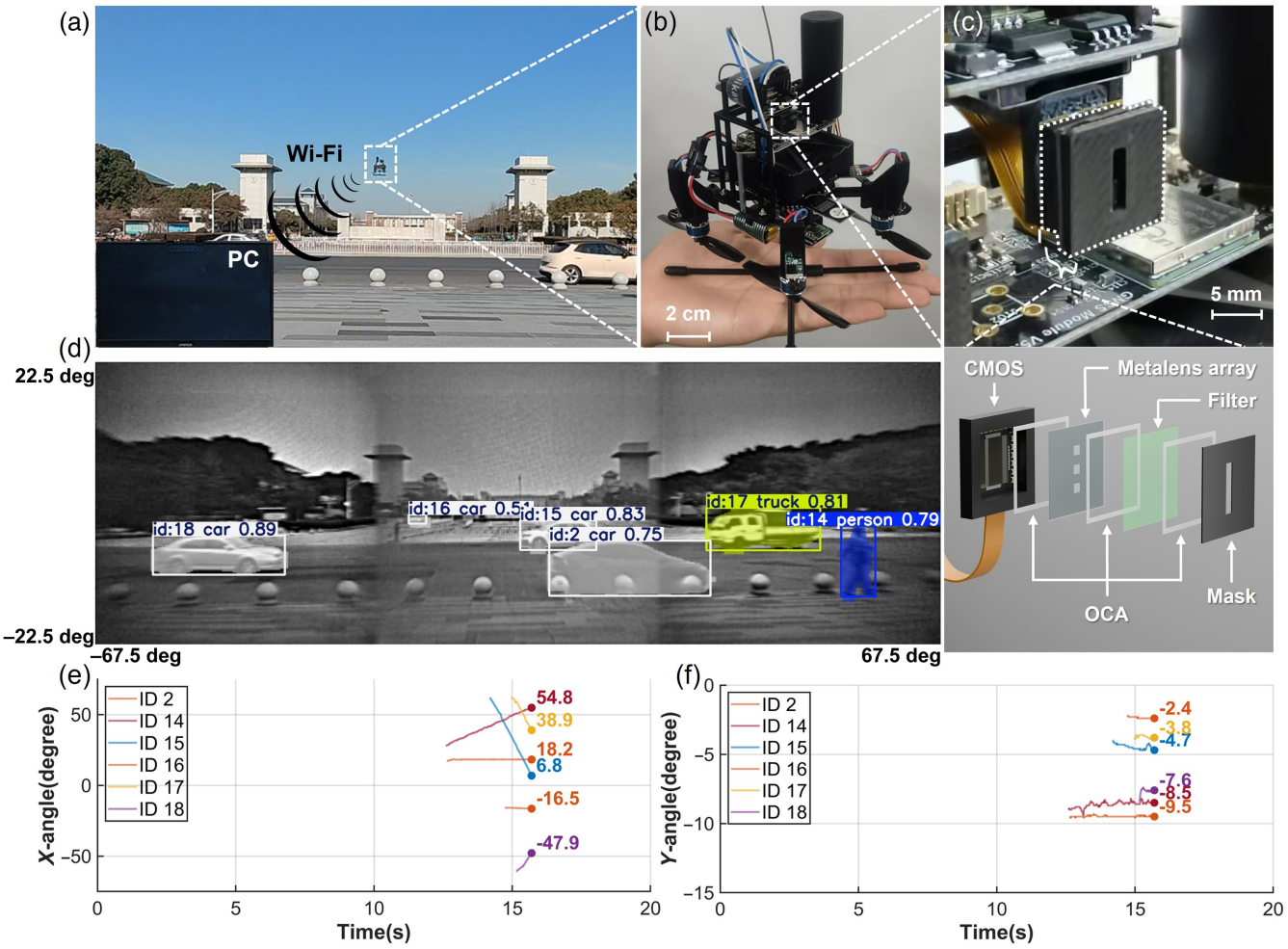


Fig. 5 High-mobility detection of real dynamic objects. (a) Schematic of high-mobility detection based on an NNOD-integrated miniature unmanned aerial vehicle (MAV). Images captured by the NNOD are wirelessly transmitted to a laptop through WiFi signals. (b) Photograph of the MAV integrated with NNOD. The MAV is 8 cm × 8 cm × 15.6 cm in size. The NNOD is integrated on top of the 3D-printed bracket. (c) Zoom-in picture and architecture of the integrated NNOD. The mask, optical filter, metalens array, and CMOS sensor are fixed together using optical clear adhesive (OCA). The distance between the metalens array and the CMOS is precisely controlled, equaling the metalens focal length. (d) One specific frame of the continuous detection results of a real wide-angle street scene. The text above each dynamic target represents the identification number, the recognized category, and the recognition confidence/accuracy. The angular localization results of each target at (e) the *x*-direction and (f) the *y*-direction are also presented. The horizontal axis represents the entire time of the continuous detection. Each curve corresponds to a target object in panel (d), showing its continuous localization results from the moment it entered the wide-angle scene to the current frame. The numbers at the end of the curves are the localization results (Video S2, MP4, 5.30 MB [URL: https://doi.org/10.1117/1.APN.5.4.046013.s2]; Video S3, MP4, 5.48 MB [URL: https://doi.org/10.1117/1.APN.5.4.046013.s3]).

is integrated at the center of the top of the 3D-printed bracket to prevent obstruction of the FOV [Fig. 5(b)]. The zoom-in picture of NNOD and its architecture are shown in Fig. 5(c), where a 3D-printed mask, a 550-nm optical filter, the metalens array, and the CMOS sensor are compactly bonded together using optical clear adhesive (OCA). By adjusting the thickness of the metalens array and CMOS sensor to match the metalens focal length, optimal imaging performance of far-field objects would be achieved.

The NNOD continuously captures real-world scenes to form a video stream, which is transmitted through the interfaces of each module at high speed and sent wirelessly via WiFi to a laptop. Through the laptop, OpenCV retrieves the video stream from the wireless receiver module into system memory, where it is subsequently processed by the neural network program for intelligent image quality enhancement and object detection. Each frame of the wide-angle scene undergoes a complete processing cycle, from acquisition and wireless transmission (less than 10 ms) to subsequent processing in a computer (110 ms for module 1 and 20 ms for module 2), within ∼130 to 140 ms. This allows the detection results to be displayed at a rate of ∼10 frames per second, achieving real-time intelligent detection of dynamic real-world environments with near-zero latency. The details of the wireless signal transmission process and the data stream processing step on the laptop are provided in Secs. S16 and S17 in the Supplementary Material, respectively.

We conducted real-time motion target detection using the NNOD-integrated MAV system in a real-world street scene, as shown in Fig. 5(a). Owing to the far-field imaging design of the metalens array, targets beyond 10 m can be effectively detected. Video S2 illustrates the NNOD-integrated MAV system's real-time target detection performance throughout the entire flight sequence, from takeoff and hovering to landing, demonstrating its flexibility and high-quality sensing capabilities. To validate the localization performance of the system in real-world scenarios, the NNOD was deployed in a stationary configuration for target detection. A specific frame of the video is extracted to show the intelligent detection performance, including the various recognition categories and high accuracy of objects in the wide-FOV scene, as shown in Fig. 5(d). It can be seen that the recognition accuracy remains unaffected even when objects are positioned at large viewing angles, at the stitching seams of sub-images, or at areas where two objects overlap. The corresponding angular positions of these objects in this video frame are extracted respectively at x and y directions [Figs. 5(e) and 5(f)]. Each curve in these two figures corresponds to a target object in the wide-angle image [Fig. 5(d)], showing its continuous localization results from the moment it entered the wide-angle scene to the current frame. The numbers at the end of the curves indicate the specific localization results. The recognition and localization results of more video frames are presented in Sec. S18 in the Supplementary Material. Video S3 shows the wide-angle imaging and intelligent detection and localization of dynamic targets over a continuous period. Although the localization error tends to increase under dynamic conditions, further improvements in the mechanical flight stability of MAVs are expected to enhance both the detection accuracy and the operational stability of the system.

3 Discussion and Conclusion

Our wide-angle metalens array design achieves an impressive horizontal FOV exceeding 135 deg, whereas the vertical imaging FOV also reaches up to 45 deg, benefiting from the ray tracing optimization. This performance far exceeds that of conventional refractive lenses with comparable sizes (see Sec. S19 in the Supplementary Material for details of performance comparison). Conventional wide-angle imaging typically requires complex fisheye lenses composed of multiple bulky refractive elements. By contrast, our compact wide-angle metalens array not only eliminates this limitation but also offers additional advantages such as ultrathin form factor,[11] compact polarization imaging,[43,44] and snapshot hyperspectral imaging[45]—capabilities inherently difficult to achieve with traditional bulky optics. Although the current horizontal FoV extension already meets the demands of most ground-based applications, such as autonomous driving or landscape surveillance, the viewing range can be further extended toward 2D coverage using a metalens array architecture by arranging sub-metalenses with distinct linear tilt phases along both axes, enabling near-omnidirectional angular observation. Furthermore, our design methodology can be readily extended to the infrared regime by replacing the visible-light material with infrared-compatible materials such as silicon and using an infrared detector. This would enable additional functionalities not accessible in the visible domain, including night vision and improved imaging through fog or smoke. Certainly, increasing the number of designed metalenses can further expand the imaging field-of-view, theoretically approaching 180 deg. However, as the viewing angle continues to grow, issues such as significant intensity attenuation and severe imaging aberrations become increasingly prominent and difficult to overcome by image processing, thereby substantially degrading the imaging quality. These challenges may be addressed by employing advanced strategies such as topology optimization to design more complex and tailored metasurface unit structures or using tilted nanopillars or substrate-tilted etching techniques[46] to enhance coupling efficiency at oblique incidence.

The detection performance of the NNOD is not only determined by the designed metalens array and the neural network, but also significantly influenced by other integrated components and the computational capabilities of the connected processing system. For example, the pixel count and shutter speed of the CMOS sensor directly impact the spatial and temporal resolution of detection, respectively—higher pixel density enables finer spatial localization, whereas faster shutter speeds improve the system's ability to capture rapid motion with greater temporal precision. Increasing the metalens aperture size or focal length could further utilize the CMOS pixel array, allowing higher resolution imaging of distant objects and finer details. In practice, the achievable frame rate is constrained by the camera sensor's resolution-dependent maximum frame rate, potential buffer stalls under WiFi congestion, on-chip processing latency, the wireless transmission bandwidth, and the computational performance of the processing computer. Higher transmission bandwidth and more powerful computing capabilities enable fast processing and recognition of high-resolution images, allowing real-time detection to be performed at higher frame rates for smoother and more responsive performance. In addition, the flight stability of the MAV is important for achieving accurate target localization and improving the NNOD's robustness to external disturbances such as wind. Meanwhile, real-time inertial data from the MAV can be incorporated to estimate platform motion and compensate for the associated image perturbations, thereby improving the robustness of target

recognition and maintaining localization accuracy under flight disturbances.

In conclusion, we developed an NNOD with wide-angle imaging, intelligent recognition, and localization capabilities. The system was integrated into an MAV, demonstrating advantages in real-time high-mobility detection and achieving high-performance operation across diverse real-world scenarios. The core component of the NNOD is a metalens array composed of three metalenses, achieving an imaging FOV exceeding 135 deg. Complemented by a comprehensive neural network, the NNOD delivers an average recognition accuracy of various objects over 96% and an angular localization precision of 0.1 deg. A detailed comparison of our NNOD with other representative object detection works is provided in Table S1 in the Supplementary Material. Our work presents significant potential for advancing the future developments of intelligent positioning, smart transportation, environmental monitoring, etc., in the emerging low-altitude economy and the next-generation mobile communication technologies.

4 Appendix: Methods

4.1 Ray Tracing Optimization

The ray tracing optimization process was conducted in Ansys Zemax OpticStudio software, aiming to obtain the phase profile of the center-FOV metalens. The optimized phase enables the metalens to achieve an imaging FOV of −22.5 to 22.5 deg. Due to the symmetrical design, we only need to analyze the light rays with incident angles within 0 to 22.5 deg. This angle range is divided into 4.5 deg intervals, resulting in light rays with six kinds of incident angles for analysis. As the rays corresponding to each incident angle must be designed to exit through filter apertures located at different positions, a multiconfiguration design strategy in Zemax is required, with each angle's rays assigned to a distinct configuration. In optimization, the objective is to minimize the average size of the focal spots for rays at the six incident angles on the focal plane, which could be expressed as

$$[\phi(r), p_1, \ldots, p_6] = \underset{\phi(r), p_1, \ldots, p_6}{\arg\min} \sum_{i=1}^{6} S_i(\phi(r), \theta_i), \tag{2}$$

where $\phi(r)$ is the metalens phase profile, p_i is the distance between the metalens and the i'th filtering aperture, θ_i is the i'th incident angle, and S_i is the focal spot size of the i'th incident light. Although only six independent angles were analyzed to obtain the optimized phase, it was sufficient to achieve high-quality focusing for any angle within the continuous range.

In Zemax, the metalens is modeled as a "Binary2" surface type, which can be designed as a pure phase plane with no thickness. The phase profile was defined as an even order polynomial of the radial coordinate r as

$$\phi(r) = \sum_{i=1}^{5} a_i \cdot \left(\frac{r}{R}\right)^{2i}, \tag{3}$$

where R is the radius of the metalens, and the coefficients a_i were optimized to minimize the focal spot sizes. A fifth-order polynomial is sufficiently accurate to represent any symmetrical phase. The positions of the filtering apertures were also set as parameters to be optimized, which indirectly determines the regions of the metalens illuminated by incident light at different angles. By constraining the operands "TTLT" and "TTGT," which represent the minimum and maximum distances between the filtering apertures and the metalens, and by adjusting the weights of these operands in the "Merit Function Editor" in Zemax, a complex optimization objective tailored to our design can be achieved. Subsequently, employing the global optimization search in Zemax enables the acquisition of the optimal phase profile for the center-FOV metalens. The final optimized coefficients of the phase polynomial are shown in Table 1.

Table 1 Optimized coefficients of the phase polynomial.

a_1	a_2	a_3	a_4	a_5
−888.0541	4.7298	1.8393	−1.3519	0.0076

4.2 Fabrication of the Metalenses

First, a 1000-nm-thick silicon nitride film was deposited on the SiO_2 substrate using Plasma-enhanced chemical vapor deposition (PECVD). To enhance the imaging quality, a shading mask was then fabricated outside the metalenses area. After preprocessing, the substrates were loaded into the electron beam lithography system (Elionix ELS-F125, Tokyo, Japan) after sequentially coated with ∼220 nm of resist film (AR-N7520) and ∼50 nm of conductive adhesive (AR-PC 5090). After exposure, the conductive polymer was dissolved in water, and the resist was developed in a resist developer solution (ZX 238) for 40 s and fixer (DI water) for 60 s in sequence. Then, the substrates were sequentially thermal evaporated with a 10-nm-thick chromium adhesion layer, a 50-nm-thick aurum shading mask, and a 30-nm-thick chromium layer as the protective coating for subsequent processing. After that, the samples were immersed in NMP solution for lift-out, leaving the shading mask adhered to the silicon nitride film. The following steps involve the fabrication of the metalens structure. A 225-nm PMMA A4 resist film and a 50-nm-thick layer of conductive polymer (AR-PC 5090) were spin-coated onto the sample in sequence and loaded into the electron beam lithography system (Elionix ELS-F125). After exposure, the conductive polymer was then dissolved in water, and the resist was developed in a resist developer solution (MIBK:IPA = 1:3) for 120 s and a fixer (IPA) for 60 s in sequence. A 40-nm electron thermal evaporated chromium layer was used to reverse the generated pattern with a liftoff process and was then used as a hard mask for dry etching the silicon layer. Finally, the sample is dry etched and is immersed in the stripping solution (ceric ammonium nitrate); the chromium layer is removed from the substrate, leaving only the desired metalens structure on the substrate. The whole fabrication process is shown in Fig. S5(b) in the Supplementary Material.

4.3 Imaging Distortion Correction

The key step in distortion correction is establishing the mapping relationship between object points and image points. By applying the inverse transformation of this mapping, distortion can be corrected. Although the mapping is initially determined during the design of the metalens array using ray tracing methods, errors in sample fabrication and testing prevent its direct application. Traditional approaches, such as the Brown–Conrady

model,[47] express the object point coordinates as even-degree polynomial functions of the image point coordinates, with the polynomial coefficients determined through sampling and fitting to establish the mapping relationship. However, these traditional methods are primarily effective for addressing center-symmetric distortion types, such as pincushion or barrel distortion. As a result, for the side-FOV metalenses in our lens array, which experience asymmetric distortion, traditional methods are not sufficient. To address this, we define the mapping relationship as a polynomial that includes both odd- and even-degree terms, shown as

$$\begin{aligned} x &= f_a(x_d, y_d) \\ &= a_1 + a_2 \cdot x_d + a_3 \cdot y_d + a_4 \cdot x_d^2 + a_5 \cdot x_d \cdot y_d + a_6 \cdot y_d^2 \\ &\quad + a_7 \cdot x_d^3 + a_8 \cdot x_d^2 \cdot y_d + a_9 \cdot x_d \cdot y_d^2 + a_{10} \cdot y_d^3 y \\ &= f_b(x_d, y_d) \\ &= b_1 + b_2 \cdot x_d + b_3 \cdot y_d + b_4 \cdot x_d^2 + b_5 \cdot x_d \cdot y_d + b_6 \cdot y_d^2 \\ &\quad + b_7 \cdot x_d^3 + b_8 \cdot x_d^2 \cdot y_d + b_9 \cdot x_d \cdot y_d^2 + b_{10} \cdot y_d^3, \end{aligned} \tag{4}$$

where x and y refer to the object points coordinates, x_d and y_d refer to the image points coordinates, and a_i and b_i are the coefficients to be determined. For both the center-FOV and side-FOV metalenses, a 9×9 uniform 2D grid of points was selected and then imaged by the metalenses. Through the fitting of these object points and image points, the polynomial coefficients can be determined. Through this polynomial, the distorted coordinates (x_d, y_d) can be substituted to calculate the undistorted coordinates (x, y), thereby achieving distortion correction [Fig. S10(a) in the Supplementary Material].

Another important issue is that if the (x_d, y_d) coordinates on the distorted image are traversed to calculate the corresponding (x, y) coordinates, it may lead to many (x, y) coordinates that are not integer values. To address this issue, we traverse the integer pixel coordinates (x, y) to find the corresponding (x_d, y_d) and then use a quadratic interpolation method to determine the intensity of pixel (x, y) based on the intensities of the four nearest integer coordinates around (x_d, y_d). According to Eq. (4), the corresponding coordinates (x_d, y_d) for each (x, y) are first determined, which are mostly noninteger pixels. The four integer points closest to (x_d, y_d) are labeled as $(x_{\text{floor}}, y_{\text{floor}})$, $(x_{\text{floor}}, y_{\text{ceil}})$, $(x_{\text{ceil}}, y_{\text{floor}})$, and $(x_{\text{ceil}}, y_{\text{ceil}})$, respectively. Then, the intensity of pixel (x, y) would be calculated as

$$\begin{aligned} Q(x, y) = (y_{\text{ceil}} - y_d) \cdot [Q_{11} \cdot (x_{\text{ceil}} - x_d) + Q_{21} \cdot (x_d - x_{\text{floor}})] \\ + (y_d - y_{\text{floor}}) \cdot [Q_{12} \cdot (x_{\text{ceil}} - x_d) + Q_{22} \cdot (x_d - x_{\text{floor}})], \end{aligned} \tag{5}$$

where Q_{11}, Q_{12}, Q_{21}, and Q_{22} represent the intensities of pixels $(x_{\text{floor}}, y_{\text{floor}})$, $(x_{\text{ceil}}, y_{\text{floor}})$, $(x_{\text{floor}}, y_{\text{ceil}})$, and $(x_{\text{ceil}}, y_{\text{ceil}})$, respectively [Fig. S10(b) in the Supplementary Material]. After sequentially determining the intensity of all the pixels (x, y), the complete distortion-corrected image would be obtained. We then imaged a chessboard pattern using the metalens array and applied the distortion correction method. The comparison between the distorted and corrected images of a chessboard, as shown in Fig. S10(c) in the Supplementary Material, highlights the high-quality distortion correction effectiveness achieved by our approach.

Disclosures

All authors declare that they have no competing interests.

Code and Data Availability

Source codes for the metalens array wide-angle imaging processing and object recognition and localization are available at: https://doi.org/10.5281/zenodo.15768045 (Ref. 48). Additional codes are available upon request from the corresponding author, Ji Chen (jichen@seu.edu.cn). The data that support the findings of this study are provided in the Supplementary Material. Source data are provided with this paper.

Contributions

J.C. developed the idea. J.C., M.L., and Y.Wu proposed the design and performed the numerical simulation. Y.Wu, M.L., Z.Y., and Z.-W.Z. conducted the optical testing experiments with the help of J.C. and Y.Wang. M.L. and Y.Z. conducted the neural network calculation with the assistance of J.C. Y.Wu., C.H., and J.J. fabricated and characterized the samples with the help of K.Q. Z.Y. and M.L. conducted the device integration under the guidance of B.Z. J.C., T.L., and Z.Z. supervised the project. J.C. analyzed the results and wrote the paper under the guidance of S.Z. All authors contributed to the discussion.

Acknowledgments

This work was supported by the National Key R&D Program of China (Grant No. 2022YFA1404301), the National High-Level Personnel of Special Support, Basic Research Program of Jiangsu (Grant No. BK20252002), the National Natural Science Foundation of China (Grant Nos. 12104223, 62325504, 61960206005, and 61871111), the Jiangsu Key R&D Program Project (Grant No. BE2023011-2), the Young Elite Scientists Sponsorship Program by CAST (Grant No. 2022QNRC001), the Fundamental Research Funds for the Central Universities (Grant Nos. 2242022R10128 and 2242022k60001), and the Project of National Mobile Communications Research Laboratory (Grant No. 2026A03).

References

1. Z. Zhao et al., "Object detection with deep learning: a review," *IEEE Trans. Neural Networks Learn. Syst.* **30**, 3212–3232 (2019).
2. S. Feng et al., "Intelligent driving intelligence test for autonomous vehicles with naturalistic and adversarial environment," *Nat. Commun.* **12**, 748 (2021).
3. Z. Zou et al., "Object detection in 20 years: a survey," *Proc. IEEE* **111**, 257–276 (2023).
4. F. Bao et al., "Heat assisted detection and ranging," *Nature* **619**, 743–748 (2023).
5. B. Li, Q. Lin, and M. Li, "Frequency angular resolving LiDAR using chip scale acousto optic beam steering," *Nature* **620**, 316–322 (2023).
6. A. Voulodimos et al., "Deep learning for computer vision: a brief review," *Comput. Intell. Neurosci.* **2018**, 7068349 (2018).
7. L. Liu et al., "Deep learning for generic object detection: a survey," *Int. J. Comput. Vis.* **128**, 261–318 (2020).
8. K. Li et al., "Object detection in optical remote sensing images: a survey and a new benchmark," *ISPRS J. Photogramm. Remote Sens.* **159**, 296–307 (2020).

9. W. Li et al., "Intelligent metasurface system for automatic tracking of moving targets and wireless communications based on computer vision," *Nat. Commun.* **14**, 989 (2023).
10. J. Pang et al., "Reconfigurable perovskite X ray detector for intelligent imaging," *Nat. Commun.* **15**, 1769 (2024).
11. M. Khorasaninejad et al., "Metalenses at visible wavelengths: diffraction limited focusing and subwavelength resolution imaging," *Science* **352**, 1190–1194 (2016).
12. S. Wang et al., "Broadband achromatic optical metasurface devices," *Nat. Commun.* **8**, 187 (2017).
13. M. Khorasaninejad et al., "Achromatic metalens over 60 nm bandwidth in the visible and metalens with reverse chromatic dispersion," *Nano Lett.* **17**, 1819–1824 (2017).
14. S. Wang et al., "A broadband achromatic metalens in the visible," *Nat. Nanotechnol.* **13**, 227–232 (2018).
15. T. Li et al., "Revolutionary meta imaging: from superlens to metalens," *Photonics Insights* **2**, R01 (2023).
16. J. Chen et al., "Metamaterials: from fundamental physics to intelligent design," *Interdiscip. Mater.* **2**, 5–29 (2023).
17. H. Pahlevaninezhad et al., "Nano optic endoscope for high resolution optical coherence tomography in vivo," *Nat. Photonics* **12**, 540–547 (2018).
18. C. Hao et al., "Compact meta camera for intelligent wide angle and low light imaging," *Laser Photonics Rev.* **2025**, e00803 (2025).
19. J. Chen et al., "China's top 10 optical breakthroughs: advancements in optical imaging devices based on metalens arrays (invited)," *Laser Optoelectron. Prog.* **61**, 2200001 (2024).
20. Y. Zhang et al., "Deep learning enhanced high quality imaging in metalens integrated camera," *Opt. Lett.* **49**, 2853–2856 (2024).
21. Y. Wu et al., "Tbps wide field parallel optical wireless communications based on a metasurface beam splitter," *Nat. Commun.* **15**, 7744 (2024).
22. M. Li et al., "Millimeter precision positioning for wide angle indoor area enabled by metalens integrated camera," *Nanophotonics* **13**, 4101–4110 (2024).
23. W. Chen et al., "A broadband achromatic metalens for focusing and imaging in the visible," *Nat. Nanotechnol.* **13**, 220–226 (2018).
24. S. Shrestha et al., "Broadband achromatic dielectric metalenses," *Light Sci. Appl.* **7**, 85 (2018).
25. R. J. Lin et al., "Achromatic metalens array for full colour light field imaging," *Nat. Nanotechnol.* **14**, 227–231 (2019).
26. Q. Fan et al., "Trilobite inspired neural nanophotonic light field camera with extreme depth of field," *Nat. Commun.* **13**, 2130 (2022).
27. A. Arbabi and A. Faraon, "Advances in optical metalenses," *Nat. Photonics* **17**, 16–25 (2023).
28. A. Arbabi et al., "Miniature optical planar camera based on a wide angle metasurface doublet corrected for monochromatic aberrations," *Nat. Commun.* **7**, 13682 (2016).
29. A. Martins et al., "On metalenses with arbitrarily wide field of view," *ACS Photonics* **7**, 2073–2079 (2020).
30. C. Hao et al., "Single layer aberration compensated flat lens for robust wide angle imaging," *Laser Photonics Rev.* **14**, 2000017 (2020).
31. J. Engelberg et al., "Near IR wide field of view Huygens metalens for outdoor imaging applications," *Nanophotonics* **9**, 361–370 (2020).
32. M. Y. Shalaginov et al., "Single element diffraction limited fisheye metalens," *Nano Lett.* **20**, 7429–7437 (2020).
33. F. Zhang et al., "Extreme angle silicon infrared optics enabled by streamlined surfaces," *Adv. Mater.* **33**, 2008157 (2021).
34. X. Luo et al., "Recent advances of wide angle metalenses: principle, design, and applications," *Nanophotonics* **11**, 1–20 (2022).
35. J. Chen et al., "Planar wide angle imaging camera enabled by metalens array," *Optica* **9**, 431–437 (2022).
36. A. J. Satti et al., "Nearly 180° field of view metalens for miniaturized imaging systems," *Laser Photonics Rev.* **2025**, e00098 (2025).
37. R. Kaur and S. Singh, "A comprehensive review of object detection with deep learning," *Digit. Signal Process.* **132**, 103812 (2023).
38. J. Ding et al., "Object detection in aerial images: a large scale benchmark and challenges," *IEEE Trans. Pattern Anal. Mach. Intell.* **44**, 7778–7796 (2022).
39. J. Du, "Understanding of object detection based on CNN family and YOLO," *J. Phys. Conf. Ser.* **1004**, 012029 (2018).
40. P. Jiang et al., "A review of YOLO algorithm developments," in *8th Int. Conf. Inf. Technol. Quant. Manag. ITQM 2020 2021 Dev. Glob. Digit. Econ. COVID 19*, Vol. **199**, pp. 1066–1073 (2022).
41. J. Terven, D. M. Córdova Esparza, and J. A. Romero González, "A comprehensive review of YOLO architectures in computer vision: from YOLOv1 to YOLOv8 and YOLO NAS," *Mach. Learn. Knowl. Extr.* **5**, 1680–1716 (2023).
42. M. Hussain, "YOLO v1 to YOLO v8, the rise of YOLO and its complementary nature toward digital manufacturing and industrial defect detection," *Machines* **11**, 677 (2023).
43. N. A. Rubin et al., "Matrix Fourier optics enables a compact full Stokes polarization camera," *Science* **365**, eaax1839 (2019).
44. F. J. Li et al., "Meta grating lens–based monolithic polarization camera," *Sci. Adv.* **11**, eadx9886 (2025).
45. X. Hua et al., "Ultra compact snapshot spectral light field imaging," *Nat. Commun.* **13**, 2732 (2022).
46. Y. Wang et al., "Metalens with tilted structures for high efficiency focusing at large angle incidences," *Chin. Opt. Lett.* **22**, 053601 (2024).
47. D. C. Brown, "Decentering distortion of lenses," *Photogramm. Eng. Remote Sens.* **32**, 444–462 (1966).
48. J. Chen, M. Li, and Z. Yuan, "Neural nanophotonic object detector with an ultra wide FOV," arXiv:2505.19379 (2025).
49. E. Tseng et al., "Neural nano optics for high quality thin lens imaging," *Nat. Commun.* **12**, 6493 (2021).
50. P. Chakravarthula et al., "Wirtinger holography for near eye displays," *ACM Trans. Graph.* **38**, 1 (2019).
51. H. Zheng et al., "Multichannel meta imagers for accelerating machine vision," *Nat. Nanotechnol.* **19**, 471–478 (2024).
52. D. G. Suárez Forero et al., "Spin selective strong light–matter coupling in a 2D hole gas–microcavity," *Nat. Phys.* **20**, 1234–1240 (2024).
53. L. Bian et al., "Towards large scale single shot millimeter wave imaging for low cost security inspection," *Nat. Commun.* **15**, 6459 (2024).
54. E. Choi et al., "360° structured light with learned metasurfaces," *Nat. Photonics* **18**, 848–855 (2024).
55. G. Kim et al., "Metasurface driven full space structured light for three dimensional imaging," *Nat. Commun.* **13**, 5920 (2022).
56. R. J. Martins et al., "Metasurface enhanced light detection and ranging technology," *Nat. Commun.* **13**, 5724 (2022).
57. J. Kang et al., "Tough interface enabled stretchable electronics using non stretchable polymer semiconductors and conductors," *Nat. Nanotechnol.* **17**, 1265–1271 (2022).
58. Y. Y. Xie et al., "Metasurface integrated vertical cavity surface emitting lasers for programmable directional lasing emissions," *Nat. Nanotechnol.* **15**, 125–130 (2020).
59. S. Q. Li et al., "Phase only transmissive spatial light modulator based on tunable dielectric metasurface," *Science* **364**, 1087–1090 (2019).
60. J. Sun et al., "Large scale nanophotonic phased array," *Nature* **493**, 195–199 (2013).
61. W. Liu et al., "Image adaptive YOLO for object detection in adverse weather conditions," *Proc. AAAI Conf. Artif. Intell.* **36**, 1792–1800 (2022).
62. Y. Song et al., "Vision transformers for single image dehazing," *IEEE Trans. Image Process.* **32**, 1927–1941 (2023).
63. G. Narasimhan and K. Nayar, "Vision and the atmosphere," *Int. J. Comput. Vis.* **48**, 233–254 (2022).

64. A. Olesel, G. Ramponi, and J. Mathews, "Image enhancement via adaptive unsharp masking," *IEEE Trans. Image Process.* **9**, 505–510 (2000).
65. C. Huang, C. Cheng, and S. Chiu, "Efficient contrast enhancement using adaptive gamma correction with weighting distribution," *IEEE Trans. Image Process.* **22**, 1032–1041 (2013).
66. J. Redmon et al., "You only look once: unified, real time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 779–788 (2016).
67. J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 7263–7271 (2017).
68. J. Redmon and A. Farhadi, "YOLOv3: an incremental improvement," arXiv:1804.02767 (2018).
69. M. Hussain, "YOLOv5, YOLOv8 and YOLOv10: the go-to detectors for real-time vision," arXiv:2407.02988 (2024).
70. Y. Wang and M. Liao, "YOLOv1 to YOLOv10: the fastest and most accurate real time object detection systems," *APSIPA Trans. Signal Inf. Process.* **13**, 1 (2024).
71. A. Bochkovskiy, C. Wang, and M. Liao, "YOLOv4: optimal speed and accuracy of object detection," arXiv:2004.10934 (2020).
72. M. Sohan, T. Sai, and V. Rami, "A review on YOLOv8 and its advancements," in *Data Intell. Cognit. Inform.*, pp. 529–545 (2024).
73. X. Li et al., "Generalized focal loss: towards efficient representation learning for dense object detection," *IEEE Trans. Pattern Anal. Mach. Intell.* **45**, 3139–3153 (2022).
74. Y. Lin et al., "Microsoft COCO: common objects in context," *Lect. Notes Comput. Sci.* **8693**, 740–755 (2014).

Ji Chen is a Research Fellow at the School of Information Science and Engineering at Southeast University. He received his PhD from Nanjing University in 2019. His research interests include metasurfaces, advanced imaging, and optical wireless communications.

Tao Li is a professor at the College of Engineering and Applied Sciences at Nanjing University. He received his PhD from Nanjing University in 2005. His research interests include optical metamaterials, topological photonics, and on-chip photonic integrations.

Zaichen Zhang is a professor at the School of Information Science and Engineering at Southeast University. He received his PhD from Hong Kong University in 2002. His research interests include 6G mobile information and communication systems, optical wireless communication technology, industrial Internet of Things, and quantum information technology.

Biographies of the other authors are not available.