



Deep learning enhanced achromatic imaging with a singlet flat lens

SHANSHAN HU,¹ XINGJIAN XIAO,¹ XIN YE,¹ RONGTAO YU,¹
YANHAO CHU,¹ JI CHEN,² SHINING ZHU,¹ AND TAO LI^{1,*} 

¹National Laboratory of Solid State Microstructures, Key Laboratory of Intelligent Optical Sensing and Manipulation, Jiangsu Key Laboratory of Artificial Functional Materials, School of Physics, College of Engineering and Applied Sciences, Nanjing University, Nanjing 210093, China

²National Mobile Communications Research Laboratory, School of Information Science and Engineering, Frontiers Science Center for Mobile Information Communication and Security, Southeast University, Nanjing 211189, China

*taoli@nju.edu.cn

Abstract: Correction of chromatic aberration is an important issue in color imaging and display. However, realizing broadband achromatic imaging by a singlet lens with high comprehensive performance still remains challenging, though many achromatic flat lenses have been reported recently. Here, we propose a deep-learning-enhanced singlet planar imaging system, implemented by a 3 mm-diameter achromatic flat lens, to achieve relatively high-quality achromatic imaging in the visible. By utilizing a multi-scale convolutional neural network (CNN) imposed to an achromatic multi-level diffractive lens (AMDL), the white light imaging qualities are significantly improved in both indoor and outdoor scenarios. Our experiments are fulfilled via a large paired imaging dataset with respect to a 3 mm-diameter AMDL, which guaranteed with achromatism in a broad wavelength range (400–1100 nm) but a relative low efficiency (~45%). After our CNN enhancement, the imaging qualities are improved by ~2 dB, showing competitive achromatic and high-quality imaging with a singlet lens for practical applications.

© 2023 Optica Publishing Group under the terms of the [Optica Open Access Publishing Agreement](#)

1. Introduction

Broadband achromatic imaging plays a significant role in various applications, such as mobile photography, projectors, telescopes, and virtual and augmented reality. To correct the chromatic aberration in traditional refractive lenses, multiple lenses with different material dispersions are combined to a lens group in an imaging module [1]. However, the realization of achromatism increases the volume, weight, complexity, and cost of the optical system. This is contrary to the trend of future development toward integration and miniaturization of optical components. Flat lenses, including metalenses and diffractive lenses, provide a new strategy for lightweight and compact optical imaging systems. Metalens is a metasurface with a focusing phase [2], which has the advantages of ultra-thin, ultra-light, and flexibly design. To walk towards the practical applications, researchers have made great efforts to improve imaging performance of metalenses [3,4], such as improved efficiency [5,6], large numerical aperture [7–10], achromatism [11–20], wide-field metalenses [21–29], and so on. Although many works have reported achromatic imaging with a relatively broad bandwidth, the very limited lens thickness restricts its ability to compensate for chromatic aberration in large lens sizes. As a consequence, the limited lens size and low working efficiency fail to meet the demand for wide applications with high-quality imaging [30–32]. As a paralleled branch of flat lenses, multi-level diffractive lenses (MDLs) composed of wavelength-level diffractive rings can be designed for achromatism with carefully designed heights of rings [33–36]. Many significant functionalities have been implemented with MDLs, like achromatism [32,34,36–38], large numerical aperture [39], super-resolution [40], extended depth of focus [41], large field of view (FOV) [42] and so on. However, the

reported AMDLs are difficult to achieve a large diameter, high NA, high focus efficiency and small thickness simultaneously, though they have made progresses over the metalenses.

Very recently, our group proposed a new framework to achieve large-diameter achromatic flat lens based on light frequency-domain coherence design and structural optimization [32]. Based on this method, we have successfully designed and fabricated several achromatic multi-level lenses (AMDLs) with diameters ranging from 1 to 10 mm. These lenses operate in a super broadband from 400 to 1100 nm with relatively good comprehensive performance. However, the focusing efficiency of AMDLs is also not high enough ($<70\%$) even for the smallest one, which will inevitably lead to background noises. Therefore, it is very important to improve the image quality of AMDLs in real-world landscape imaging. In recent years, deep learning is becoming the research hotspots of artificial intelligence and has achieved excellent performance in the field of imaging processing tasks such as image deblurring, denoising, super-resolution, and so on [43]. Many researchers have successfully migrated this technique to imaging techniques from intelligent design to subsequent processing, bringing new possibilities to improve the imaging performance of optical systems [42,44–51]. Considering that further optimization of AMDL in structural design and fabrication scheme is facing great challenges in the short term, it is feasible and effective to improve the AMDL system performance by incorporating deep learning method.

In this work, we propose a deep-learning-enhanced planar camera assembled with a singlet AMDL, which enables high-quality achromatic imaging in the visible. To efficiently incorporate the neural network and the AMDL imaging system, we experimentally acquire the point spread function (PSF) images of our system and perform an accurate imaging simulation based on the imaging formation model [42,52]. Then, we obtain a large paired dataset for training the network model without tedious manual image collection and complexed image registration. The network model here is the multi-scale CNN. Finally, we compare and analyze the image performance with and without neural network processing, which shows significant improvement of the image quality of indoor and outdoor scenes acquired by AMDL camera with the neural network processing.

2. AMDL imaging system and imaging performance

Figure 1(a) shows the schematic of an AMDL. Briefly, we maximize the frequency domain coherence function at the focus of the lens to realize broadband achromatism, which is realized by optimizing the height distribution along the radial axis through genetic algorithm, Hooke-Jeeves algorithm and gradient descent. The details of the design process are shown in the previous work [32]. As a result, an AMDL with diameter equal to 3 mm and average focusing efficiency about 40% is demonstrated theoretically and experimentally. The profile of the AMDL is shown in Fig. 1(b). The measured focal lengths and focus efficiencies are displayed in Fig. 1(c), and more details can be found in Ref. [32]. To investigate the imaging performance of the AMDL, we mounted the AMDL to the complementary metal-oxide semiconductor (CMOS) through an adjustable lens barrel to construct an AMDL camera (see Fig. 1(d)). The total size of the camera is 4.2 cm \times 4.2 cm \times 1.9 cm. This AMDL camera is employed for an indoor imaging of objects of color checkers, stamp, ruler, with imaging distance of about 2 m, and an outdoor imaging for various features and depths, as depicted in Fig. 1(e). Although the imaging results exhibit no chromatic aberration, there are apparent background noises and low signal-to-noise ratio of both indoor and outdoor images.

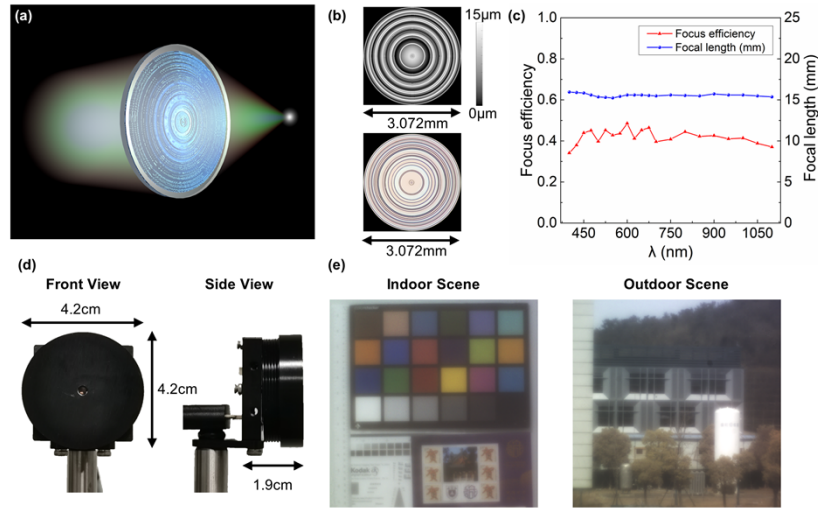


Fig. 1. Illustration of AMDL imaging system. (a) Schematic of the AMDL. (b) The profile and photographs of fabricated AMDL. (c) The measured focus efficiencies and focal lengths of the AMDL. (d) Physical diagram of the AMDL integrated device. The left shows the front view, and the right shows the side view. (e) The imaging results for real-world scenes. The left column shows the indoor scene and the right column shows the outdoor scene.

3. Multi-scale CNN architecture

Imaging degradation of the AMDL camera is mainly induced by the relatively low focusing efficiency of the AMDL and the non-uniform blurs and noise during the imaging process. Inspired by the deep learning technique for imaging enhancement, we utilize a multi-scale CNN to learn the transformation from degraded images to ground truths in a semi-blind manner. The multi-scale architecture that integrates the features in different scales, is widely adopted in low-level vision tasks [52]. As shown in Fig. 2, the network composes of three parallel branches, learning coarse level, middle level and fine level information, respectively. Specifically, the network takes the blurry image as the input of the fine-scale branch and utilizes the average pooling layer to generate the down-sampled images in the form of Gaussian Pyramids as the input of the middle-scale branch and the coarse-scale branch. The scale ratio between consecutive branches is 0.5. For the structure of the branch, there is a convolutional layer at first, stacking 10 residual blocks without batch normalization layers and rectified linear unit before the block output, which is denoted as ResBlocks. Then, following the last convolution layer to generate a corresponding sharp result. Before the last convolutional layer of the coarse scale and middle scale branch, feature maps are conveyed to the next finer branch. To match the input size of the next finer branch, the output feature maps pass an upsampling layer and concatenate with the finer scale input, then fused by a purple convolution layer. Finally, the output of the fine-scale branch is the recovered sharp result for an enhanced image.

The loss function is combined with multi-scale content loss and perceptual loss [46,52]. The multi-scale content loss is the weighted sum of the mean square error (MSE) between the blurry input and the sharp output of all branches. The perceptual loss is calculated in the original scale branch, using the first 25 layers of VGG19 [53]. The total loss function is defined as follows

$$L_{total} = \sum_{k=1}^3 w_k \|R_k - S_k\|^2 + \lambda \|\varphi(R_1) - \varphi(S_1)\|_2^2 \quad (1)$$

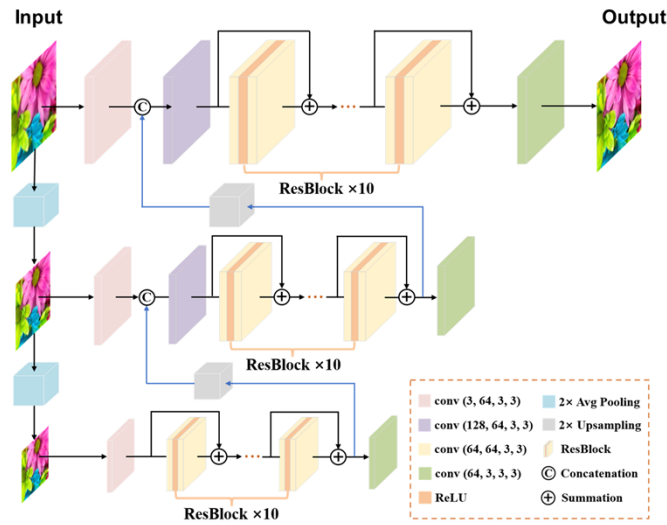


Fig. 2. Architecture of the deep multi-scale CNN. The network contains three branches primarily composed of modified residual blocks. Using blurred images as input to the network, each branch processes features with different scales. During the training process, the output of each branch is hierarchically fused into the previous branch in a bottom-up manner. The original scale image is the final sharp result.

where R_k , S_k , and w_k denote the output, ground truth, and weight constant at scale level k . φ represents the operation of VGG19. Here, the λ is the weight constant of perceptual loss. We minimize the total loss function value to optimize the network parameters in training.

4. Experiment

4.1. Imaging formation model

In supervised learning, it is critical to get the corresponding ground truths of the degraded images captured with the proposed optical imaging system. The performance of numerous models trained on synthetic data would degrade significantly when applied to real-world data. Therefore, acquiring paired real-world image datasets or datasets as close as possible to reality is crucial to ensure the restoration ability of the model on degraded real-world images. Manual acquisition requires not only a heavy shooting workload but also complex alignment which influences the restoration quality. Although there are some mechanical acquisition manners, they require complex mechanical systems additionally to apply in various scenes [42]. By comparing with synthetic images degraded with a Gaussian degenerate kernel, the image simulation method is more accurate and general for different optical systems. This method is based on the imaging formation model, which is defined as [42,52]:

$$I(x, y) = S(x, y) * P(x, y) + n(x, y) \quad (2)$$

where $I(x, y)$, $S(x, y)$, and $P(x, y)$ are the degraded image, latent sharp ground truth, and the PSF respectively. The PSF represents the response of the optical imaging system to a point source, which is spatially various with different spatial coordinates (x, y) . $n(x, y)$ is the noise during imaging, and $*$ is the operation of convolution.

There are two stages for simulating modern digital imaging. First, using the ray tracing or fast Fourier transform (FFT) methods to compute the optical PSF with the parameters of the

optical lens, and then introducing the sensor and the imaging signal processing (ISP) parameters for the image simulation [54,55]. To minimize the difference between simulated data and real data, we adopt the imaging simulation method. Unlike the full computational simulation, we obtain the real PSFs images of the proposed AMDL camera in experiment and convolve them with sharp ground truth images to generate a large amount of paired image datasets quickly and accurately, without elaborating manual acquisition and complicated computational simulation. In the following, the experiment of acquiring the original PSF images and the generation of the datasets are interpreted in detail.

4.2. PSF dataset acquisition

As shown in Fig. 3(a), a PSF images acquisition experimental setup is built up. The point light source is composed of a white LED light source (440-670 nm) and a 100 μm pinhole assembled on the LED through adapter elements. The point light source is located on motorized stages, which are stacked in XY configuration with the travel range of 300 mm. The AMDL camera is placed on the optical rail with manual stages to horizontal and vertical displacement. The object distance is set to be 90 cm, which is approximated as infinity imaging. In the experiment, we control the point light source move to the center of the x -axis of the motorized stages and define this location as the center of imaging, and adjust the position of the camera so that the point light source is imaged in the center of CMOS. Then, we move the point light source to the original location and use the motorized stages to control the precise movement of the point light source along the X and Y direction, and the movement step is set to be 1 cm. The AMDL camera shoots the images of the point light source at different locations sequentially, which are the spatially varied PSF original images. In the total field of view, 427 images are captured, some of which are shown in Fig. 3(b).

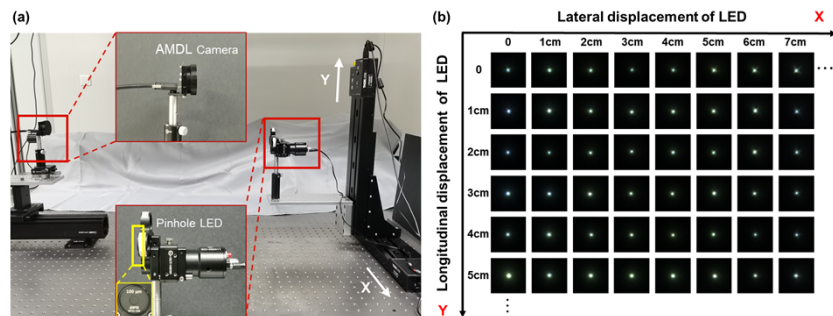


Fig. 3. Illustration of the acquisition process for the PSF dataset. (a) Schematic of the experimental setup. It consists of the motorized stages, the white LED light source, the pinhole which size is 100 μm , and the AMDL integrated device, and the white LED and pinhole are equivalent to a point light source. (b) The partial imaging results of PSF at different locations. Moving the point light source laterally and longitudinally and capturing the imaging results. The step length is 1 cm.

Considering the rotation invariance of the lens and the experimental error, we rotate, scale, and exchange the channel of the original PSF images (Size: 51×51) randomly to augment the dataset for preventing overfitting and improving generalization performance. Further, to improve the robustness of the model to different noise levels, random Gaussian noises with the standard deviation from 0 to 5 are added. The sharp ground truth dataset we used is picked out of 10000 images from the public Coco dataset. We randomly crop the ground truth to 224×224 patches and convolute them with a PSF image of the augmented dataset. Repeating this step, we obtain the simulated images of our imaging system and paired sharp ground truths to train the model.

4.3. Neural network training details

After the acquisition and generation of datasets, we set the weight constant w_k in Eq. (1) to be $w_k = 1, 0.7,$ and 0.5 respectively when $k = 1, 2, 3$. λ is set to 0.002 . We utilize the Adam [56] optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The learning rate is initialized as 0.0001 for the first 25 epoch and decreases linearly to $1/10$ for another 25 epoch. The training epoch and batch size is 50 and 4 respectively. The model is implemented using PyTorch framework. All training experiments are run on a desktop with Nvidia Titan V GPU. The training time using a single GPU is about 8 hours.

5. Results

5.1. Quantitative analysis of simulation imaging results

The network performance is evaluated quantitatively on the imaging simulated images. As shown in Fig. 4, we select four different scenes to simulate, compare and analyze. The number of PSFs array is 30×24 in the selected simulated region of FOV. To cover spatially variant degradation and as close as possible to real imaging, the sharp ground truths are split into 30×24 image patches with overlap and the size is 224×224 . Then, the sharp patches are convolved with the PSF of the corresponding position respectively to get the degraded image patches. We stitch and fusion the degraded image patches to obtain the blurry images, which are the second column in Fig. 4. Then the degraded image patches are put into the model, and the output patches are stitched into the recovered results. As shown in Fig. 4, regions of the center or edge of FOV are highlighted to compare each pair of the ground truths, the blurry input, and the network output. From the recovered fine features in highlighted regions, the image quality of the network outputs is improved significantly compared with the corresponding blurry image. We calculate the averaged peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) of each image pair to evaluate the recovery performance of the model quantitatively. The averaged PSNR of network outputs improve by 2 dB approximately on the selected scenes in Fig. 4, and the improved averaged SSIM also proves the improved performance of the model.

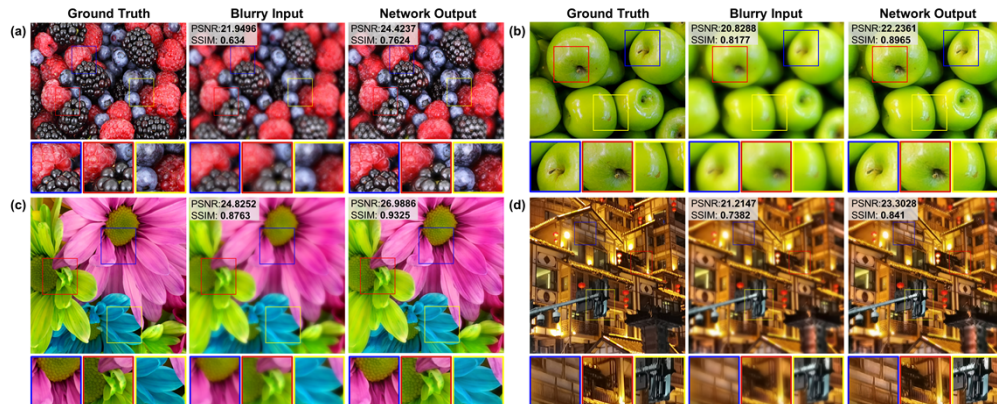


Fig. 4. The simulated results of the AMDL imaging system. The first column presents the ground truths of different scenes, the second column presents their corresponding simulated imaging results of the AMDL imaging system, and the third column presents their corresponding network outputs. The blue/red/yellow highlighted regions denote different regions on images. The averaged PSNR and SSIM are located in the top-left corner of the images.

We also conduct ablation experiments to compare the effectiveness of the network branch number. In Table. 1, the 1-layer model represents the fine-scale branch is preserved and the

2-layer model has two branches without the coarse-scale branch. We test these models on the images in Fig. 4 and calculate the PSNR and SSIM. The average PSNR and SSIM of the 3-layer model are better than other models.

Table 1. Ablation study on the number of the network branch

Scene	Blurry Image		Deep Multi-scale Convolutional Neural Network					
	PSNR	SSIM	1 layer		2 layers		3 layers	
Fig. 4(a)	21.9496	0.634	24.1926	0.7627	24.3627	0.76	24.4237	0.7624
Fig. 4(b)	20.8288	0.8177	21.2682	0.8922	21.4989	0.8918	22.2361	0.8965
Fig. 4(c)	24.8252	0.8763	26.0795	0.9322	26.3083	0.934	26.9886	0.9325
Fig. 4(d)	21.2147	0.7382	23.1121	0.839	23.1989	0.8393	23.3028	0.841
Avg	22.2046	0.7666	23.6631	0.8565	23.8422	0.8563	24.2378	0.8581

5.2. Qualitative analysis of experimental imaging results

To validate the reconstruction performance of the trained model for realistic photography, we use the experimental imaging results captured by our AMDL camera as the input of the trained model. The original degraded images and the corresponding network output are displayed in Fig. 5.

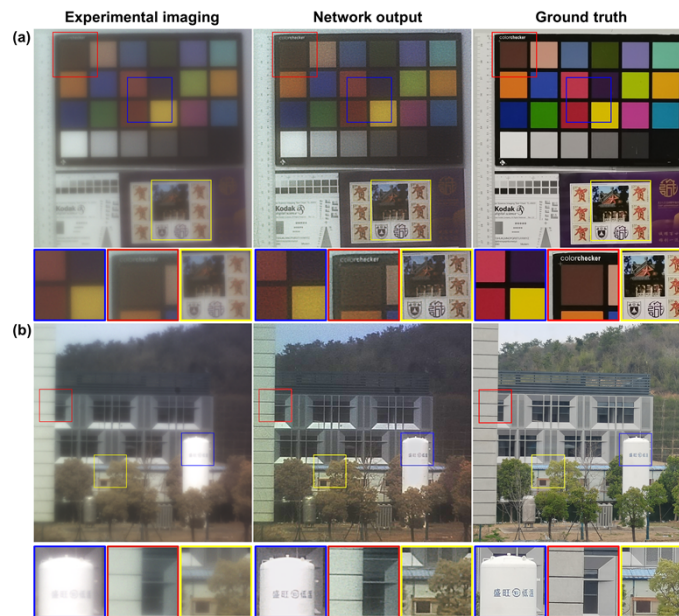


Fig. 5. The experimental results of the AMDL imaging system. The first column presents the experimental imaging results of the indoor scene (the first row) and outdoor scene (the second row). The second column presents the network output results of their corresponding blurry imaging results. The third column presents the corresponding photographs of the cell phone cameras.

In the highlight regions of the indoor scene, the edge and boundaries of objects, the detailed graphics and text are recovered to clarity. The building surface, the windows, the trees, and the text on the white liquid tank with different depths in the outdoor scene are also sharper and clearer. The reconstruction results demonstrate that the trained model can remove the degradation caused by the blurs and noise effectively and improve imaging performance for various scenes

with different characteristics and depths. The right column in Fig. 5 displays the ground truth of the indoor and outdoor scenes, which are captured by a commercial cell phone (Huawei nova 7 pro). It is observed the resolution of the Network output images are almost comparable with the ground truth, though the comprehensive imaging quality is still inferior to Ground truth. Note that we capture the PSF at a fixed distance, while the trained model is quite applicable for outdoor imaging. It is due to the fixed distance in PSF acquisition is already approximately regarded as imaging at infinity. Moreover, the scaling/rotation operation on original PSF images during augmentation improves the robustness and universality of the model and endows the model with the ability to deal with various image degradation.

6. Conclusion

In conclusion, we have successfully employed the CNN method to enhance the achromatic imaging in the visible based on a singlet AMDL. The methods for experimentally acquiring PSFs and generating large datasets by optical imaging simulation reduce the differences of simulation and real data, ensuring the reconstructed performance of the model to improve the image quality. Our result suggests the possibility of high-quality imaging based on a singlet achromatic flat lens within a compact camera prototype, and inspires further planar optical devices for practical applications. It is hopeful that the imaging performance of the AMDL can be further improved by correcting for higher order aberrations with advanced computational techniques, including topological optimization, inverse design, deep-learning-based co-optimization. In addition to the optimization of lens design, the development of manufacturing techniques also provides a lot of room for further improvement of AMDL performance.

Funding. National Key Research and Development Program of China (2022YFA1404301); National Natural Science Foundation of China (12174186, 62288101, 62325504, 92250304).

Acknowledgments. Tao Li thanks the support from Dengfeng Project B of Nanjing University.

Disclosures. The authors declare no conflicts of interest.

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

References

1. F. L. Pedrotti, L. M. Pedrotti, and L. S. Pedrotti, *Introduction to Optics* (Cambridge University, 2017).
2. N. Yu, P. Genevet, and M. A. Kats, *et al.*, "Light propagation with phase discontinuities: generalized laws of reflection and refraction," *Science*. **334**(6054), 333–337 (2011).
3. T. Li, C. Chen, and X. Xiao, *et al.*, "Revolutionary meta-imaging: from superlens to metalens," *Photon. Insights* **2**(1), R01 (2023).
4. M. K. Chen, Y. Wu, and L. Feng, *et al.*, "Principles, functions, and applications of optical meta-lens," *Adv. Opt. Mater.* **9**(4), 2001414 (2021).
5. M. Khorasaninejad, W. T. Chen, and R. C. Devlin, *et al.*, "Metalenses at visible wavelengths: Diffraction-limited focusing and subwavelength resolution imaging," *Science*. **352**(6290), 1190–1194 (2016).
6. A. Arbabi, Y. Horie, and A. J. Ball, *et al.*, "Subwavelength-thick lenses with high numerical apertures and large efficiency based on high-contrast transmitarrays," *Nat. Commun.* **6**(1), 7069–7074 (2015).
7. Z.-B. Fan, Z.-K. Shao, and M.-Y. Xie, *et al.*, "Silicon nitride metalenses for close-to-one numerical aperture and wide-angle visible imaging," *Phys. Rev. Applied* **10**(1), 014005 (2018).
8. H. Liang, Q. Lin, and X. Xie, *et al.*, "Ultrahigh numerical aperture metalens at visible wavelengths," *Nano Lett.* **18**(7), 4460–4466 (2018).
9. W. T. Chen, A. Y. Zhu, and M. Khorasaninejad, *et al.*, "Immersion meta-lenses at visible wavelengths for nanoscale imaging," *Nano Lett.* **17**(5), 3188–3194 (2017).
10. R. Paniagua-Dominguez, Y. F. Yu, and E. Khaidarov, *et al.*, "A metalens with a near-unity numerical aperture," *Nano Lett.* **18**(3), 2124–2132 (2018).
11. M. Khorasaninejad, F. Aieta, and P. Kanhaiya, *et al.*, "Achromatic metasurface lens at telecommunication wavelengths," *Nano Lett.* **15**(8), 5358–5362 (2015).
12. D. Lin, A. L. Holsteen, and E. Maguid, *et al.*, "Photonic multitasking interleaved Si nanoantenna phased array," *Nano Lett.* **16**(12), 7671–7676 (2016).
13. O. Avayu, E. Almeida, and Y. Prior, *et al.*, "Composite functional metasurfaces for multispectral achromatic optics," *Nat. Commun.* **8**(1), 14992–14998 (2017).

14. H. Li, X. Xiao, and B. Fang, *et al.*, "Bandpass-filter-integrated multiwavelength achromatic metalens," *Photonics Res.* **9**(7), 1384–1390 (2021).
15. M. Khorasaninejad, Z. Shi, and A. Y. Zhu, *et al.*, "Achromatic metalens over 60 nm bandwidth in the visible and metalens with reverse chromatic dispersion," *Nano Lett.* **17**(3), 1819–1824 (2017).
16. E. Arbabi, A. Arbabi, and S. M. Kamali, *et al.*, "Controlling the sign of chromatic dispersion in diffractive optics with dielectric metasurfaces," *Optica* **4**(6), 625–632 (2017).
17. S. Wang, P. C. Wu, and V. Su, *et al.*, "Broadband achromatic optical metasurface devices," *Nat. Commun.* **8**(1), 187–195 (2017).
18. S. Wang, P. C. Wu, and V. Su, *et al.*, "A broadband achromatic metalens in the visible," *Nat. Nanotechnol.* **13**(3), 227–232 (2018).
19. W. T. Chen, A. Y. Zhu, and V. Sanjeev, *et al.*, "A broadband achromatic metalens for focusing and imaging in the visible," *Nat. Nanotechnol.* **13**(3), 220–226 (2018).
20. S. Shrestha, A. C. Overvig, and M. Lu, *et al.*, "Broadband achromatic dielectric metalenses," *Light: Sci. Appl.* **7**(1), 85–95 (2018).
21. A. Arbabi, E. Arbabi, and S. M. Kamali, *et al.*, "Miniature optical planar camera based on a wide-angle metasurface doublet corrected for monochromatic aberrations," *Nat. Commun.* **7**(1), 13682–13690 (2016).
22. B. Groever, W. T. Chen, and F. Capasso, "Meta-lens doublet in the visible region," *Nano Lett.* **17**(8), 4902–4907 (2017).
23. Z. Lin, B. Groever, and F. Capasso, *et al.*, "Topology-optimized multilayered metaoptics," *Phys. Rev. Applied* **9**(4), 044030 (2018).
24. M. Y. Shalaginov, S. An, and F. Yang, *et al.*, "Single-element diffraction-limited fisheye metalens," *Nano Lett.* **20**(10), 7429–7437 (2020).
25. A. Martins, K. Li, and J. Li, *et al.*, "On metalenses with arbitrarily wide field of view," *Acs Photonics* **7**(8), 2073–2079 (2020).
26. F. Zhang, M. Pu, and X. Li, *et al.*, "Extreme-Angle Silicon Infrared Optics Enabled by Streamlined Surfaces," *Adv. Mater.* **33**(11), 2008157 (2021).
27. B. Xu, H. Li, and S. Gao, *et al.*, "Metalens-integrated compact imaging devices for wide-field microscopy," *Adv. Photonics* **2**(06), 066004 (2020).
28. X. Ye, X. Qian, and Y. Chen, *et al.*, "Chip-scale metalens microscope for wide-field and depth-of-field imaging," *Adv. Photonics* **4**(04), 46006 (2022).
29. J. Chen, X. Ye, and S. Gao, *et al.*, "Planar wide-angle-imaging camera enabled by metalens array," *Optica* **9**(4), 431–437 (2022).
30. F. Presutti and F. Monticone, "Focusing on bandwidth: achromatic metalens limits," *Optica* **7**(6), 624–631 (2020).
31. J. Engelberg and U. Levy, "Achromatic flat lens performance limits," *Optica* **8**(6), 834–845 (2021).
32. X. Xiao, Y. Zhao, and X. Ye, *et al.*, "Large-scale achromatic flat lens by light frequency-domain coherence optimization," *Light: Sci. Appl.* **11**(1), 323 (2022).
33. G. J. Swanson, *Binary Optics Technology: The Theory and Design of Multi-Level Diffractive Optical Elements* (MASSACHUSETTS INST OF TECH LEXINGTON LINCOLN LAB, 1989).
34. M. Meem, A. Majumder, and R. Menon, "Full-color video and still imaging using two flat lenses," *Opt. Express* **26**(21), 26866–26871 (2018).
35. L. L. Doskolovich, R. V. Skidanov, and E. A. Bezus, *et al.*, "Design of diffractive lenses operating at several wavelengths," *Opt. Express* **28**(8), 11705–11720 (2020).
36. M. Meem, A. Majumder, and S. Banerji, *et al.*, "Imaging from the visible to the longwave infrared wavelengths via an inverse-designed flat lens," *Opt. Express* **29**(13), 20715–20723 (2021).
37. P. Wang, N. Mohammad, and R. Menon, "Chromatic-aberration-corrected diffractive lenses for ultra-broadband focusing," *Sci. Rep.* **6**(1), 21545–21551 (2016).
38. N. Mohammad, M. Meem, and B. Shen, *et al.*, "Broadband imaging with one planar diffractive lens," *Sci. Rep.* **8**(1), 2799–2804 (2018).
39. M. Meem, S. Banerji, and C. Pies, *et al.*, "Large-area, high-numerical-aperture multi-level diffractive lens via inverse design," *Optica* **7**(3), 252–253 (2020).
40. S. Banerji, M. Meem, and A. Majumder, *et al.*, "Super-resolution imaging with an achromatic multi-level diffractive microlens array," *Opt. Lett.* **45**(22), 6158–6161 (2020).
41. S. Banerji, M. Meem, and A. Majumder, *et al.*, "Extreme-depth-of-focus imaging with a flat lens," *Optica* **7**(3), 214–217 (2020).
42. Y. Peng, Q. Sun, and X. Dun, *et al.*, "Learned large field-of-view imaging with thin-plate optics," *ACM Trans. Graph.* **38**(6), 1–14 (2019).
43. K. Singh, A. Seth, and H. S. Sandhu, *et al.*, "A comprehensive review of convolutional neural network based image enhancement techniques," in *2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN)* (IEEE, 2019), pp. 1–6.
44. J. Chen, S. Hu, and S. Zhu, *et al.*, "Metamaterials: from fundamental physics to intelligent design," *Interdiscip. Mater.* **2**(1), 5–29 (2023).
45. E. Tseng, S. Colburn, and J. Whitehead, *et al.*, "Neural nano-optics for high-quality thin lens imaging," *Nat. Commun.* **12**(1), 6493–6499 (2021).

46. Q. Fan, W. Xu, and X. Hu, *et al.*, “Trilobite-inspired neural nanophotonic light-field camera with extreme depth-of-field,” *Nat. Commun.* **13**(1), 2130–2139 (2022).
47. M. K. Chen, X. Liu, and Y. Sun, *et al.*, “Artificial Intelligence in Meta-optics,” *Chem. Rev.* **122**(19), 15356–15413 (2022).
48. Y. Peng, Q. Fu, and F. Heide, *et al.*, “The diffractive achromat full spectrum computational imaging with diffractive optics,” in *SIGGRAPH ASIA 2016 Virtual Reality Meets Physical Reality: Modelling and Simulating Virtual Humans and Environments* (Association for Computing Machinery, 2016), pp. 1–2.
49. M. K. Chen, X. Liu, and Y. Wu, *et al.*, “A Meta-Device for Intelligent Depth Perception,” *Adv. Mater.* **35**, 2107465 (2023).
50. X. Liu, M. K. Chen, and C. H. Chu, *et al.*, “Underwater Binocular Meta-lens,” *ACS Photonics* **10**(7), 2382–2389 (2023).
51. Y. Fan, M. K. Chen, and M. Qiu, *et al.*, “Experimental Demonstration of Genetic Algorithm Based Metalens Design for Generating Side-Lobe-Suppressed, Large Depth-of-Focus Light Sheet,” *Laser Photon. Rev.* **16**(2), 2100425 (2022).
52. S. Nah, T. Hyun Kim, and K. Mu Lee, “Deep multi-scale convolutional neural network for dynamic scene deblurring,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017), pp. 3883–3891.
53. K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv*, arXiv:1409.1556 (2014).
54. S. Chen, H. Feng, and D. Pan, *et al.*, “Optical aberrations correction in postprocessing using imaging simulation,” *ACM Trans. Graph.* **40**(5), 1–15 (2021).
55. S. Cui, B. Wang, and Q. Zheng, “Neural invertible variable-degree optical aberrations correction,” *Opt. Express* **31**(9), 13585–13600 (2023).
56. D. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” *Comput. Sci.* (2014).