

# Deep-learning-driven end-to-end metalens imaging

Joonhyuk Seo,<sup>a,†</sup> Jaegang Jo,<sup>b,†</sup> Joocheon Kim,<sup>c,†</sup> Joonho Kang,<sup>d,†</sup> Chanik Kang,<sup>a</sup> Seong-Won Moon,<sup>c</sup> Eunji Lee,<sup>e</sup> Jehyeong Hong,<sup>a,b,d</sup> Junsuk Rho,<sup>c,e,f,g,h,\*</sup> and Haejun Chung<sup>a,b,d,\*</sup>

<sup>a</sup>Hanyang University, Department of Artificial Intelligence, Seoul, Republic of Korea

<sup>b</sup>Hanyang University, Department of Electronic Engineering, Seoul, Republic of Korea

<sup>c</sup>Pohang University of Science and Technology, Department of Mechanical Engineering, Pohang, Republic of Korea

<sup>d</sup>Hanyang University, Department of Artificial Intelligence Semiconductor Engineering, Seoul, Republic of Korea

<sup>e</sup>Pohang University of Science and Technology, Department of Chemical Engineering, Pohang, Republic of Korea

<sup>f</sup>Pohang University of Science and Technology, Department of Electrical Engineering, Pohang, Republic of Korea

<sup>g</sup>POSCO-POSTECH-RIST Convergence Research Center for Flat Optics and Metaphotonics, Pohang, Republic of Korea

<sup>h</sup>National Institute of Nanomaterials Technology, Pohang, Republic of Korea

**Abstract.** Recent advances in metasurface lenses (metalenses) have shown great potential for opening a new era in compact imaging, photography, light detection, and ranging (LiDAR) and virtual reality/augmented reality applications. However, the fundamental trade-off between broadband focusing efficiency and operating bandwidth limits the performance of broadband metalenses, resulting in chromatic aberration, angular aberration, and a relatively low efficiency. A deep-learning-based image restoration framework is proposed to overcome these limitations and realize end-to-end metalens imaging, thereby achieving aberration-free full-color imaging for mass-produced metalenses with 10 mm diameter. Neural-network-assisted metalens imaging achieved a high resolution comparable to that of the ground truth image.

Keywords: visible metalens; deep learning; image restoration; full-color imaging.

Received Jun. 13, 2024; revised manuscript received Sep. 26, 2024; accepted for publication Oct. 14, 2024; published online Nov. 14, 2024.

© The Authors. Published by SPIE and CLP under a Creative Commons Attribution 4.0 International License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI.

[DOI: [10.1117/1.AP.6.6.066002](https://doi.org/10.1117/1.AP.6.6.066002)]

## 1 Introduction

The unyielding pursuit of miniaturization and performance enhancement in optical imaging systems has led to the exploration of innovative technologies beyond conventional geometric lens-based systems. While foundational to modern optics, these systems face inherent limitations, such as chromatic<sup>1,2</sup> and spherical aberrations,<sup>3,4</sup> shadowing effects,<sup>3,4</sup> bulkiness,<sup>5,6</sup> and high manufacturing costs.<sup>7–12</sup> The quest to transcend these barriers has catalyzed the advent of metalenses, a groundbreaking development poised to redefine the landscape of optical engineering.

Metalenses, characterized by their ultrathin films with meticulously arranged subwavelength structures called meta-atoms interspersed throughout, emerged as a revolutionary alternative to

overcome the drawbacks of conventional lenses. In a recent study, deep-ultraviolet immersion lithography was combined with wafer-scale nano-imprint lithography to mass-produce low-cost and high-throughput large-aperture metalenses, contributing to their commercialization.<sup>13</sup> This novel class of lenses also promises to rectify the aforementioned issues existent in conventional optics and opens a new era of compact, efficient imaging systems.<sup>1,14</sup> Central to the appeal of metalenses is their ability to serve as optimal substitutes for traditional optical elements and thereby revolutionize a broad spectrum of applications. This encompasses not only the enhancement of capabilities of optical sensors,<sup>15</sup> smartphone cameras,<sup>16,17</sup> and unmanned aerial vehicle optics<sup>18–20</sup> but also the transformation of user experiences facilitated by augmented and virtual reality devices.<sup>21,22</sup> The potential of unparalleled diffraction-limited focusing within an ultra-light and ultra-compact form factor, even in high-NA regimes,<sup>23</sup> a feat unattainable by traditional components, is the key attribute contributing to these advancements.

\*Address all correspondence to Junsuk Rho, [jsrho@postech.ac.kr](mailto:jsrho@postech.ac.kr); Haejun Chung, [haejun@hanyang.ac.kr](mailto:haejun@hanyang.ac.kr)

<sup>†</sup>These authors contributed equally to this work.

Despite these strides, the pursuit of broadband metalenses uncovers a multifaceted trade-off among focusing efficiency, lens diameter, and spectral bandwidth,<sup>24,25</sup> with the last significantly affected by chromatic aberration.<sup>26,27</sup> This interplay highlights the inherent complexity in optimizing these lenses, where improvements in one aspect may lead to compromises in others. In addition, meta-atom-based metalenses exhibit a narrow field of view (FoV) stemming from angular dispersion inherent in meta-atom-based designs.<sup>28</sup> Consequently, at present, reported broadband metalenses exhibit chromatic aberration<sup>4,27</sup> or low focusing efficiency over a large bandwidth,<sup>1,6</sup> which impedes the commercialization of metalens-based compact imaging. This compromise, by rendering the attainment of high-efficiency broadband focusing alongside minimal chromatic and angular aberration a considerable challenge, substantially restricts the performance and the range of potential applications of metalenses. Even for the ideal metalens, it may not simultaneously satisfy broadband operation and large diameter due to the physical upper bounds.<sup>24</sup> Moreover, the limitations inherent in conventional design approaches complicate efforts to effectively address these challenges in metalens development.

Recent advancements in planar lens technology have significantly improved the control of chromatic aberration, a critical factor in full-color imaging. The technique of frequency-synthesized phase engineering, proposed by Zhang et al.,<sup>29</sup> uses cascaded cholesteric liquid crystal layers to achieve RGB achromatic focusing. However, while this approach is promising, especially in the context of single-focal-plane focusing of RGB light, it does not address the challenges of scalability and mass production for practical applications.

In direct response to these challenges, we introduce an innovative, deep-learning-powered, end-to-end integrated imaging system. By synergizing a specially designed large-area mass-produced metalens<sup>13</sup> with a customized image restoration framework, we propose a comprehensive imaging solution poised to supplant conventional geometric lens-based systems. The proposed system not only effectively addresses the aberrations mentioned above but also leverages the inherent strengths of large-area mass produced metalenses to make a significant step toward high-quality, aberration-free images. Moreover, our approach distinguishes itself by suggesting a metalens image restoration framework that can fit any metalenses suffering from aberrations or low efficiency. Also, assuming the uniform quality of mass-produced metalenses, the optimized restoration model can be applied to other metalenses manufactured with the same process. The proposed imaging system may pave the way for the next generation of compact, efficient, and commercially viable imaging systems.

Other recent studies have also explored novel methodologies to address chromatic aberration and other optical challenges. Tseng et al.<sup>30</sup> developed a neural nano-optics system that integrates meta-optical design with deep learning to enhance image reconstruction. Their fully differentiable framework optimizes both the physical design of the metalens and the accompanying image processing algorithms, demonstrating significant improvements in field-of-view and color consistency. Similarly, Maman et al.<sup>31</sup> and Dong et al.<sup>32</sup> employed hyperboloid metalenses combined with deep learning to achieve RGB achromatic imaging, offering detailed insights into aberration correction and optical performance. These studies mark substantial progress over traditional achromatic lens designs, advancing the field of chromatic aberration correction. A recent study<sup>33</sup>

also introduced an end-to-end metalens design approach facilitated by computational postprocessing, offering valuable insights into the integration of learning processes with metalens design methodologies.

In contrast to these recent approaches, our system leverages a mass-produced metalens while incorporating a deep-learning-based image restoration framework, offering a scalable and high-performance solution for full-color imaging. By compensating for aberrations and efficiency loss, our system ensures broader applicability across various imaging applications. Furthermore, our approach demonstrates a unique advantage through the use of position embedding techniques, enabling the restoration of highly blurred images caused by complex aberrations. This positions our work as a significant advancement over existing solutions, with the potential to revolutionize optical imaging technologies.

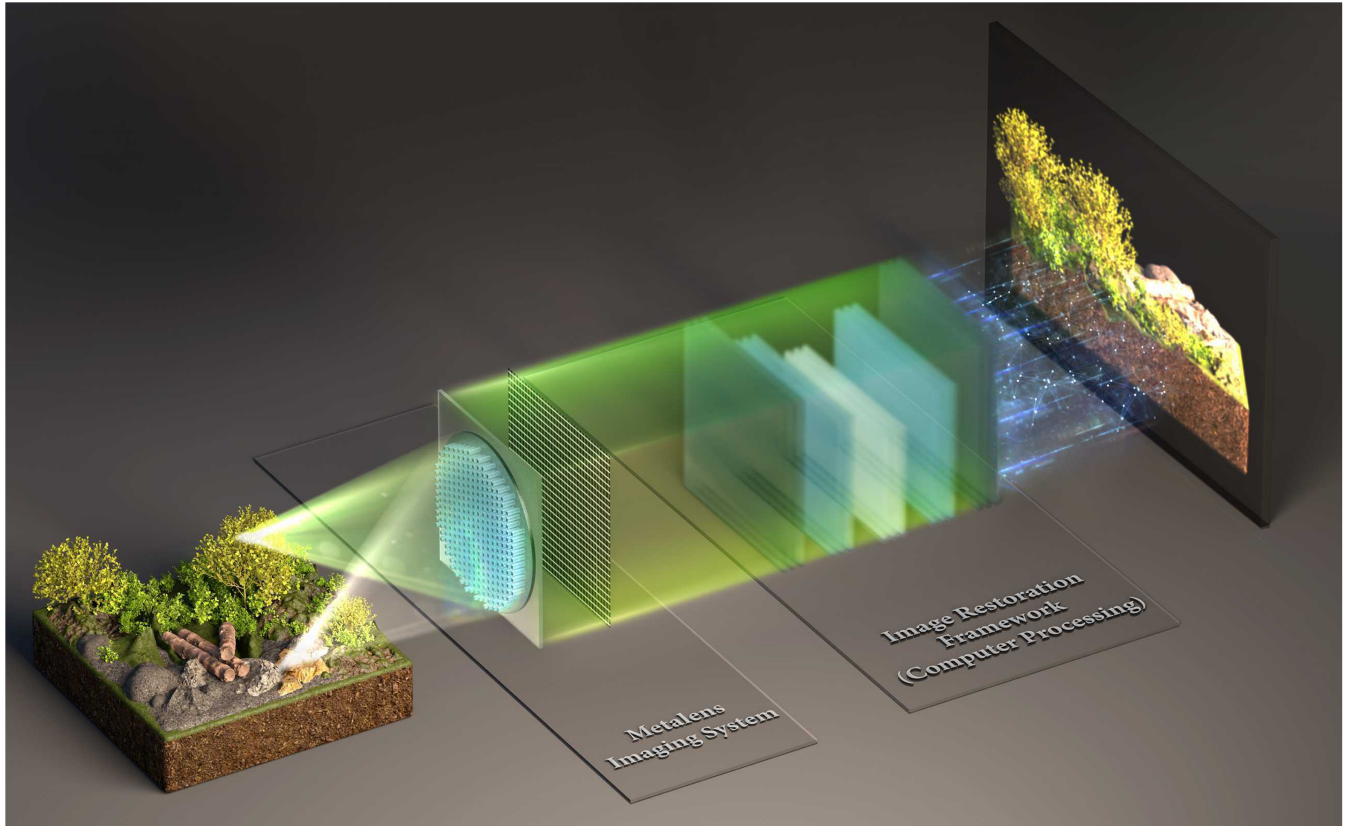
In summary, this work propels metalens technology to new heights and underscores the transformative potential of deep learning in initiating a paradigm shift in optical imaging. Through our end-to-end imaging framework, we not only demonstrate a viable pathway to surmount traditional optical limitations but also pave the way for a novel era in compact and efficient imaging solutions. This breakthrough has the potential to revolutionize the field of optical engineering, sparking new avenues of research and innovation.

## 2 Methods

A schematic of our end-to-end integrated imaging system is shown in Fig. 1. This system combines a metalens-based imaging system and a subsequent image restoration framework. The former component is tasked with acquiring the image, whereas the latter is responsible for restoring the captured image. When tailored to restore the image produced by the metalens imaging system automatically, the framework can independently generate an output image that closely approximates the quality of the ground truth image.

The metalens designed in this work is composed of an array of nanostructures with arbitrary rotational angles, with the class of metalenses designed this way being known as the Pancharatnam–Berry (PB) phase-based metalens. Despite the ability of these PB-phase-based metalenses to achieve diffraction-limited focusing,<sup>5,13</sup> they are not without their challenges. The dispersion of the meta-atoms can induce chromatic aberration, a characteristic similarly observed in diffractive lenses.<sup>26</sup> Substantial efforts have been made to achieve achromatic metalenses through dispersion engineering of meta-atoms,<sup>1,6</sup> adjoint optimization,<sup>34,35</sup> and many other methods.<sup>36,37</sup> However, the resulting metalenses still suffer from relatively low efficiency compared to single-frequency metalenses. Also, PB-phase-based metalenses are concurrently susceptible to angular aberration that originates both from Seidel aberrations<sup>3</sup> and angular dispersion of the meta-atom.<sup>28</sup> The combination of these factors sets our full color high resolution imaging apart from conventional restoration tasks,<sup>38,39</sup> thereby significantly complicating the task of restoring images captured by the metalens to their original state. Our framework thus addresses and rectifies the aberration issues of the metalens using a customized deep learning approach.

Specifically, prior to training, we gathered hundreds of aberrant images captured by the metalens imaging system, which we refer to as “metalens images.” Metalens images, which exhibit the physical defects of the metalens, were then used to train the



**Fig. 1** Schematic of our metalens imaging system.

image restoration framework. The result is a significant enhancement in the quality of the image produced by the compact metalens imaging system. The framework employed in this process is composed of two primary stages. In the first stage, the framework is optimized to reduce the discrepancy between the outputs of its restoration model and the ground truth images. Following this, an adversarial learning scheme that incorporates an auxiliary discriminator is utilized to augment the image restoration model's ability to recover lost information.

By concatenating our restoration framework to our imaging system comprising our mass-produced metalens, we construct an integrated imaging system that delivers high-quality compact imaging. This system is scalable to larger apertures and different wavelengths, thereby offering an optimal solution for a novel miniaturized imaging scheme. Importantly, the reproducibility of both the imaging system and the restoration framework not only enhances the commercial viability of this integrated system but also suggests that the commercial application of metalenses could become a reality in the near future. In the following sections, we elaborate on the construction of the integrated system, starting from the metalens to the image restoration framework.

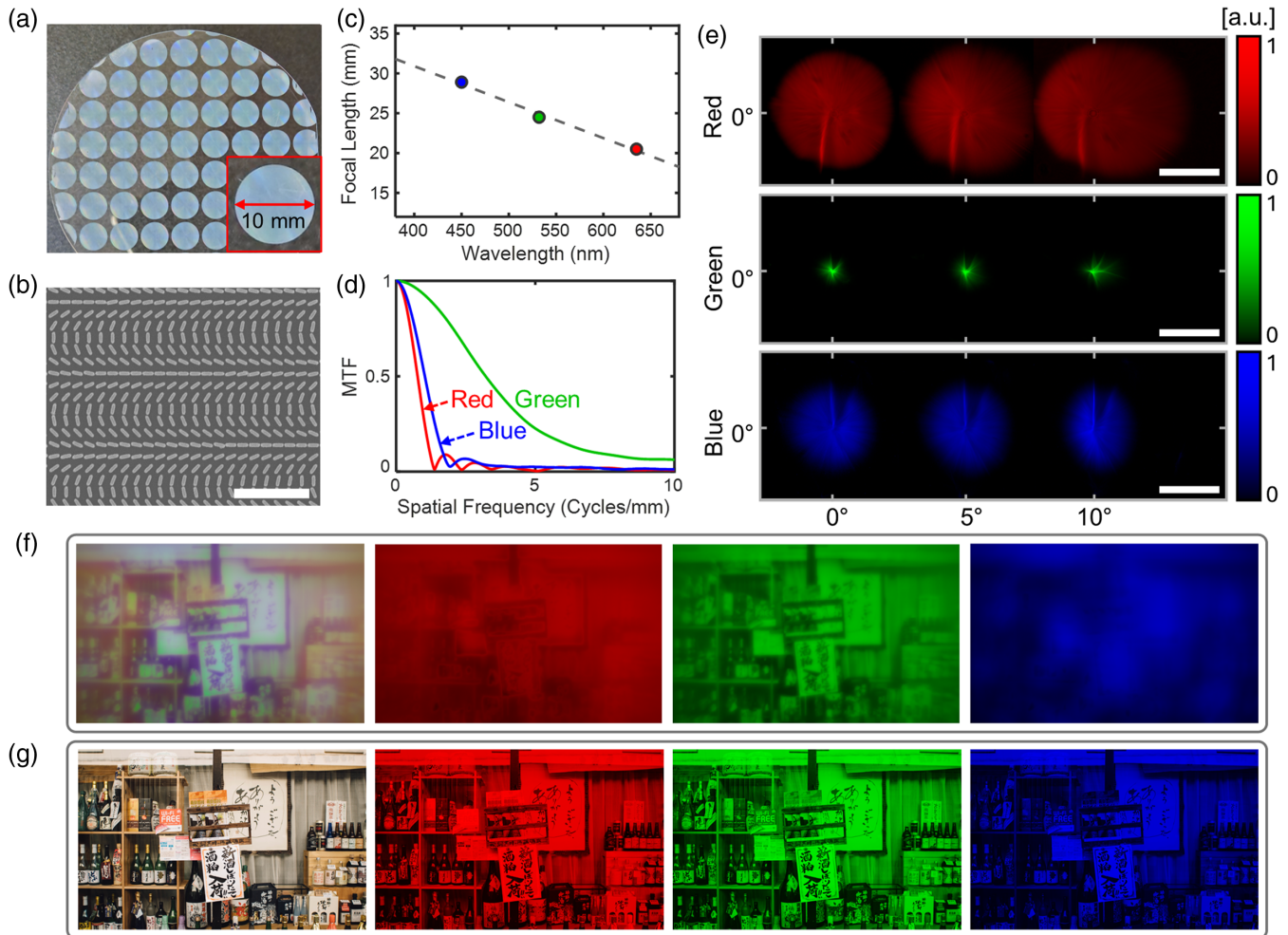
## 2.1 Metalens Imaging System

Metalenses are fabricated through nanoimprint lithography and subsequent atomic layer deposition.<sup>13</sup> Nanoimprint lithography provides the benefits of low-cost mass production and uniformity of the products.<sup>7,8,13,40</sup> Thus, we use imprinted metalenses to broadly impact our work on the commercialization of the

deep neural network (DNN)-based metalens imaging system. Figure 2(a) shows mass-produced 10-mm-diameter metalenses fabricated by nanoimprint lithography and subsequent thin-film deposition of  $\text{TiO}_2$ ; the details of the fabrication process are given in [Supplementary Material](#). As shown in Fig. 2(b), our metalens comprises nano-slabs with arbitrary rotational angles as a PB-phase-based metalens. It has a relatively high efficiency of 55.6% at 532 nm wavelength but exhibits severe chromatic aberration. As shown in Fig. 2(c), the focal lengths at wavelengths of 450, 532, and 635 nm are 29.0, 24.5, and 20.5 mm, respectively.<sup>13</sup> This wavelength-dependent focal length results in transverse axial chromatic aberration (TAC), which can be expressed as

$$\text{TAC} = |f - f_0| \frac{D}{2f}, \quad (1)$$

where  $f$  is the focal length of the incident light with a single wavelength,  $f_0$  is the distance between the metalens and image sensor, and  $D$  is the diameter of the metalens. When  $f$  differs from  $f_0$ , the incident light forms a top-hat-like point spread function (PSF) profile with a radius equal to that of TAC.<sup>26</sup> Given this, the reduction in the overall TAC in the visible band is important for high-quality metalens imaging because the blur of the image intensifies proportionally with the increase in TAC. Thus, we chose  $f_0$  to focus a green light to minimize the overall TAC. When we focus the green light (532 nm) from a far field, the red (635 nm) and blue (450 nm) lights are more defocused than the green light with TACs of 0.98 and 0.78 mm, respectively.



**Fig. 2** (a) Photograph of fabricated mass-produced 10-mm-diameter metalenses on 4" glass wafer. The inset in the red box shows enlarged image of the metalens. (b) Scanning electron microscopy (SEM) image showing the meta-atoms that compose the metalens. The scale bar is 3  $\mu\text{m}$ . (c) Focal lengths of the metalens for wavelengths of 450 nm (blue), 532 nm (green), and 635 nm (red). The dashed line indicates the linear fitting result. (d) MTFs of red, green, and blue lights with zero viewing angle. (e) PSFs of red, green, and blue lights with various viewing angles (0 deg, 5 deg, 10 deg). The scale bar is 1 mm, which indicates a distance on the image sensor. (f) Metalens image (left) and its subset images showing red, green, and blue color channels. (g) Corresponding ground truth image (left) and its subset images showing red, green, and blue color channels.

The metalens imaging system is affected by chromatic and angular aberrations and its surface defects by incomplete fabrication. To quantify these effects, we measured the PSFs and calculated the modulation transfer function (MTF) from the measured PSFs. The PSF, which is the two-dimensional (2D) intensity distribution obtained in response to a single point light source,<sup>41</sup> is a critical metric for evaluating the quality of an imaging system because it is directly related to image formation.<sup>42</sup> The MTF, calculated using the measured PSFs, describes the imaging quality in terms of resolution and contrast.<sup>41</sup> We measured the PSFs by capturing the images of collimated beams from red, green, and blue light-emitting diodes (LEDs) using the metalens imaging system and subsequently calculated the MTFs from the PSFs. The PSF measurement and the imaging setup for it are shown in Fig. S2 in the [Supplementary Material](#) and elaborated upon, and the MTF

calculation method is also subsequently explained in detail in the [Supplementary Material](#).

Figure 2(e) shows the PSFs of red, green, and blue LEDs at various viewing angles (0 deg, 5 deg, 10 deg). The PSF profiles of red and green LEDs show wide disk shapes, whereas the profile of the green LED shows irregular spark shape, implying the effect of the TAC. Thus, as shown in Fig. 2(d), the MTFs of the red and blue lights are severely lower than the MTF of the green light at all spatial frequencies. Furthermore, the PSF profiles at 0 deg viewing angle show non-ideal and circularly asymmetric shapes, which can be attributed to the defects of the metalens due to the imperfect fabrication. The PSF profiles also change their shapes with the viewing angle due to the angular aberrations, including Seidel aberrations<sup>3</sup> and the angular dispersion of the meta-atoms.<sup>28</sup> The PSF profiles of the red and green LEDs stretch to the horizontal direction as the viewing angle increases,

where the profile of the blue light shrinks. In addition, the non-uniformity of the metasurface during the fabrication process<sup>13</sup> may result in the PSFs with complex profiles not matching the PSFs from the Rayleigh-Sommerfeld diffraction formula.<sup>43</sup> As a result, the combination of these generates complex PSF profiles varying with the viewing angle and further complicates the image restoration tasks.

The effects of chromatic and angular aberrations on the metalens images can be shown by comparing them against the ground truth image. Figures 2(f) and 2(g) show the metalens image, the corresponding ground truth image, and the subset images depicting the red, green, and blue color channels. The red and blue channels of the metalens image are severely blurred from the TAC, making it difficult to recognize any objects. In addition, unlike the PSF measurements, the blue color channel appears more blurry than the red color channel due to the optical setup for data acquisitions as shown in Fig. S1 in the [Supplementary Material](#). The green channel of the metalens image shows a relatively higher resolution at the center, which gradually decreases as the viewing angle increases (e.g., the outer region of the image) due to the angular aberrations at the higher viewing angle.

## 2.2 Image Restoration Network

Computational image restorations have emerged as a prevalent approach for the enhancement of non-ideal images, such as those that are noisy<sup>44</sup> or blurred.<sup>45</sup> Classical image restoration methods achieve higher resolution by relying on linear deconvolution methods, such as applying the Wiener filter.<sup>46</sup> Deconvolution, an inverse of the convolution operation, facilitates the recovery of the original image from an image convolved with a PSF. The performance of the deconvolution process depends on two factors: the space invariance of PSF across the FoV and the low condition number for the inverse of the PSF.<sup>47</sup> However, Wiener filters exhibit limited restoration quality for imaging systems with PSFs that vary depending on the viewing angle, such as metalens imaging systems<sup>36</sup> and under-display cameras.<sup>42</sup>

An alternative restoration approach is the utilization of DNN-based image restoration. DNN-based restoration models<sup>38,39</sup> have shown superior performance compared to traditional approaches in specialized tasks, such as denoising,<sup>44</sup> de-blurring,<sup>45</sup> super-resolution,<sup>48</sup> and light-enhancement.<sup>49</sup> Furthermore, they are applicable to imaging systems with complex and combined degradations, such as under-display cameras<sup>42</sup> and the 360 deg FoV panoramic camera.<sup>50</sup> However, conventional DNN approaches are incapable of learning position-variant image degradations (e.g., position-dependent aberration of the metalens) because these methods train models with randomly cropped patches from full-resolution images, leading to the complete loss of position-dependent information.

In response to these challenges, we propose an end-to-end image restoration framework specifically tailored for the metalens imaging system to address non-uniform aberration over the wavelength and viewing angle. Contrary to the images that are subjected to restoration in typical image restoration tasks,<sup>51,52</sup> our metalens images exhibit more intense blur and significant color distortion. Consequently, the restoration of metalens images constitutes a severely ill-posed inverse problem. To address this critically underconstrained problem, we employ strong regularization. That is, we model the traits and patterns of sharp

data, performing adversarial learning in Fourier space to train the data distribution. Therefore, the restoration model  $f(y)$  is trained by minimizing

$$L(x, y, f) = E(x, f(y)) + \lambda \Phi(f(y)), \quad (2)$$

where  $E(x, f(y))$  is the image fidelity term that approximates the restored metalens image  $f(y)$  by the ground truth image  $x$ , and  $\Phi(f(y))$  is the regularization term that limits the space of  $f(y)$ . Subsequently, we apply positional embedding to learn the angular aberration of metalens imaging. Because the proposed method utilizes information on the absolute coordinates of randomly cropped patches, the model effectively trains the highly space-variant degradations.

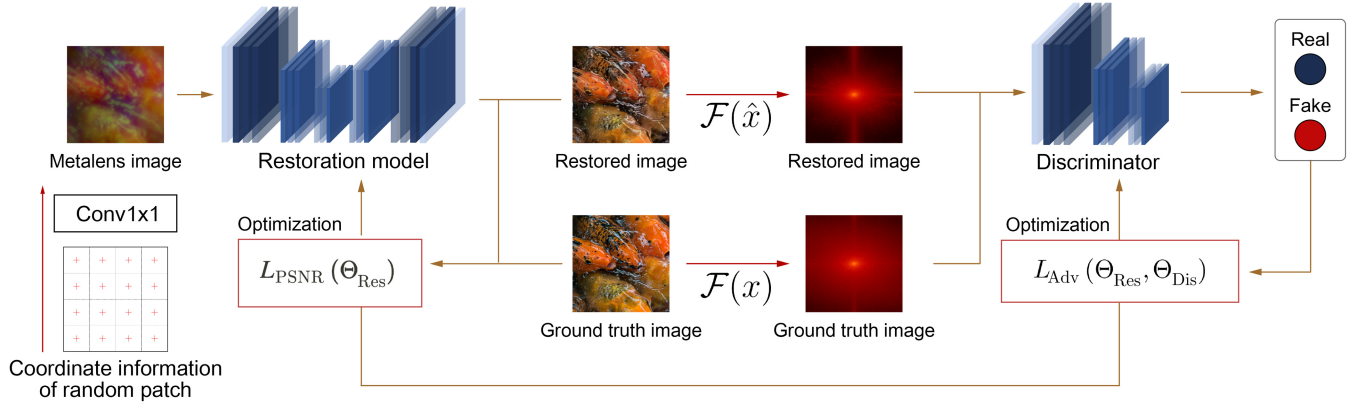
### 2.2.1 Network architecture

The architecture of our image restoration framework is depicted in Fig. 3. Our framework incorporates existing DNN architecture with our proposed methods. The training phase involved the utilization of patches randomly cropped from images at their full resolution, specifically  $1280 \times 800$  in this study. Subsequently, in the inference phase, the analysis was conducted on the entire images at their original resolution of  $1280 \times 800$ . However, we observed a statistical disagreement in the inference process of the full-resolution image, as shown in Fig. S3 in the [Supplementary Material](#). To overcome this inconsistency, we apply test-time local converter (TLC)<sup>53</sup> in the test phase, which yields a significant performance improvement. The detailed results are presented in Table S1 in the [Supplementary Material](#).

The metalens used in our study exhibits intense chromatic and angular aberrations, resulting in severe information loss in the images captured with it. Therefore, we trained the model according to the traits and patterns found in the underlying clean images to efficiently restore a wide range of spatial-frequencies and constrain the space of the latent ground truth images. Because generative models can learn complex, high-dimensional data distributions from a given dataset,<sup>54</sup> we utilized an adversarial learning scheme, one of the generative learning methods, to learn effectively the distribution of latent sharp images by introducing an auxiliary discriminator. We initially applied adversarial learning in the RGB space but observed that conspicuous pattern artifacts appeared in both the RGB and Fourier spaces (Fig. S4 in the [Supplementary Material](#)). These artifacts, related to periodic patterns, are more clearly visible in the Fourier domain than in the RGB space due to their deep connection with spectral components [Figs. S4(c) and S4(d) in the [Supplementary Material](#)]. Since the Fourier space provides a more explicit representation of these spectral components, it allowed us to identify better and address the source of the artifacts. Therefore, we transformed the data from each RGB channel into the Fourier space for adversarial learning. These Fourier space data are then used as input for the discriminator.

The training loss is composed of two distinct terms: peak signal-to-noise ratio (PSNR) loss  $L_{\text{PSNR}}$  between the reconstructed image  $\hat{x}$  and the ground truth image  $x$  and adversarial loss  $L_a$  between  $\mathcal{F}(\hat{x})$  and  $\mathcal{F}(x)$ . The PSNR loss signifies the image fidelity loss, and the adversarial loss indicates the prior regularization loss. Therefore, the total loss function  $L_{\text{Total}}$  is

$$L_{\text{Total}} = L_{\text{PSNR}} + \lambda L_a, \quad (3)$$



**Fig. 3** Proposed image restoration framework. The framework consists of an image restoration model and applies random cropping and position embedding to the input data using coordinate information of the cropped patches. To address the underconstrained problem of restoring degraded images to latent sharp images, adversarial learning in the frequency domain is applied through the FFT ( $\mathcal{F}$ ).  $\hat{x}$  and  $x$  denote the reconstructed and ground truth image, respectively. The details of the framework are in Sec. 2.

where  $\lambda$  is a hyperparameter for balancing  $L_{\text{PSNR}}$  and  $L_a$ .  $L_{\text{PSNR}}$  is calculated as follows:

$$L_{\text{PSNR}}(\hat{x}, x) = -10 \log \frac{R^2}{\text{MSE}(\hat{x}, x)}, \quad (4)$$

where  $\hat{x}$ ,  $x$ , and  $R$  denote the reconstructed image, ground truth image, and the maximum signal value of the ground truth image, respectively. MSE is the distance between the reconstructed and ground truth images and is formulated as  $\text{MSE}(\hat{x}, x) = \frac{1}{N} \sum_{n=1}^N (\hat{x}_n - x_n)^2$ .

For adversarial learning, we constructed an additional discriminator and applied spectral normalization<sup>55</sup> for training stability. In addition, we employ the GAN training scheme based on hinge loss<sup>56</sup> for enhanced stability of adversarial training. The adversarial loss  $L_a$  of the discriminator (D) and the image restoration model (G) is

$$L_a^D = \mathbb{E}_x[\max(0, 1 - D(\mathcal{F}(x)))] + \mathbb{E}_{\hat{x}}[\max(0, 1 + D(\mathcal{F}(\hat{x})))], \quad (5)$$

$$L_a^G = -\mathbb{E}_{\hat{x}}[D(\mathcal{F}(\hat{x}))], \quad (6)$$

where  $\mathcal{F}$  refers to fast Fourier transform (FFT). Here,  $\mathbb{E}_x[\cdot]$  and  $\mathbb{E}_{\hat{x}}[\cdot]$  are operators that denote the calculation of the mean of the ground truth and reconstructed images in the given minibatch, respectively. The image restoration model and discriminator each try to minimize  $L_a^G$  and  $L_a^D$ , respectively.

Degradation in the outer region of the metalens image is more pronounced than in the central region due to the angular aberration. This observation suggests that positional information is integral for understanding the degradation of the metalens imaging system. However, the training method makes it impossible for the model to learn positional information because our framework learns through random patches during training and restores full-resolution images during inference.

To address this problem, we take the coordinate values of each pixel of the patches, based on the coordinates of a full

resolution image, and map them through a  $1 \times 1$  convolutional layer. This process transforms them into proper space when generating random patches for a full-resolution image. The processed coordinate information is concatenated with metalens images corresponding to the information. The resulting concatenated data are used as input data. This approach enables the model to learn and leverage positional information effectively, thereby enhancing its performance in restoring full-resolution images.

### 2.2.2 Data acquisition

The training data for the metalens imaging system were obtained by capturing ground truth images displayed on the 85" monitor (Fig. S1 in the [Supplementary Material](#)). For training, we utilized the DIV2K dataset.<sup>51</sup> This dataset contains 2K resolution images of various objects, thereby providing environmental diversity. The ground truth images for training were obtained by cropping the center of the dataset images by  $1280 \times 800$  resolution to ensure that the ground truth images fit within the FoV of the metalens imaging system.

The positions of the objects in both the metalens image and the corresponding ground truth image were matched for effective training. Raw metalens images with  $5472 \times 3648$  resolution were rotated, cropped to  $5328 \times 3328$  resolution, and resized to  $1280 \times 800$  resolution to match the corresponding ground truth images. The rotation angle and cropping parameters were optimized to maximize the structural similarity measure (SSIM) between the metalens images and the corresponding ground truth images. Finally, we divide the dataset into 628 and 70 images for training and testing, respectively.

### 2.2.3 Training details

As mentioned in network architecture section, training was conducted using patches that were randomly cropped from full resolution images. While larger receptive fields offer more comprehensive semantic information, they also increase the training time and computational complexity. Consequently, to strike a balance between performance and training duration in the proposed model, we set the patch size at  $256 \times 256$  and the batch

size at 16. In addition, transformations such as horizontal and vertical flips and transpositions were randomly applied, and then coordinate information of the patches was loaded under these configurations.

The model used in this paper can be divided into two components, the first of which is the image restoration model. The width of the starting layer of the network is set to 32, which doubles as the network delves deeper into each successive level. The encoder and decoder of the network are each composed of four levels. To address the inconsistency between training and testing, TLC is adopted during the testing phase. The numbers of input and output channels of the  $1 \times 1$  convolutional layer that processes coordinate information are both set to 2. The second part is the discriminator, where its width is set to 64, and all layers have the same width. The discriminator is composed of five blocks. Moreover, spectral normalization<sup>55</sup> is applied to stabilize the learning process. Additional information and configurations of our architectures are provided in Table S3 and Fig. S5 in the [Supplementary Material](#).

The training was executed in two stages. In the first stage, the metalens images were restored to clean images using the image restoration model, and in the second stage, adversarial learning was performed using the discriminator after expressing the restored and ground truth images in the spatial frequency domain through fast Fourier transform (FFT). Because the spatial-frequency domain data converted through FFT are complex (comprising real and imaginary parts), these parts were represented as a 2D vector. This allowed the data in the spatial-frequency domain to be expressed as real vectors, which were then used as inputs into the discriminator.

During the training process, the number of iterations was set to 300,000. In the image restoration model, AdamW was used as the optimizer with the learning rate initially set to  $3 \times 10^{-4}$  and gradually decreased to  $10^{-7}$  following the cosine annealing schedule; the betas were [0.9, 0.9]. For the discriminator optimizer, Adam was used with the learning rate set to  $3 \times 10^{-4}$ , identical to the restoration model, but the betas were set to [0.0, 0.9]; NVIDIA RTX 4090 24 GB was used as a computational resource for this training.

#### 2.2.4 Statistics details

Statistical hypothesis testing was performed using the statistical functions of the SciPy library in Python using 70 test images. Two-sided paired *t*-tests were used to compare the performances of models at the  $P < 10^{-4}$  significance level.

### 3 Results

In this study, we have introduced a deep-learning-powered, end-to-end integrated imaging system. We now assess its capability in various perspectives to restore metalens images to their clean states, addressing severe chromatic and angular aberrations inherent in our large-area mass-produced metalens. In order to draw a comparison between the images produced by our framework and those captured with the metalens, we restored a total of 70 metalens images to their undistorted state. Given these pairs of images, we conduct a thorough assessment of our system's efficacy in image restoration, employing a comprehensive set of performance metrics tailored to each category of interest under evaluation. We also compare our framework with state-of-the-art models, including restoration models for natural images (MIRNetv2,<sup>57</sup> HINet,<sup>58</sup> NAFNet<sup>38</sup>). Furthermore, we conducted training and inference on newly collected outdoor images to

verify our framework's learning capability (Figs. S7 and S8 in the [Supplementary Material](#)). Detailed information on outdoor image restoration is in the [Supplementary Material](#).

Figures 4 and S6 in the [Supplementary Material](#) comprehensively show the qualitative restoration results of our integrated imaging system by comparing the ground truth, metalens, and system outcome images. Notably, the images captured by the metalens are marred by pronounced chromatic aberrations, manifesting as a noticeable disparity in the clarity of red and blue components in comparison to green, thereby engendering significant blurring. Furthermore, this aberration is accompanied by a loss in high-frequency information, leading to the erosion of fine details present in the original images. A particularly marked manifestation of this degradation is observed in the peripheral regions (marked by a yellow box) as compared to the central zone (highlighted by a red box) in Fig. 4, where the images exhibit enhanced blurring, resulting in the obliteration of sharp details and the predominance of a specific hue.

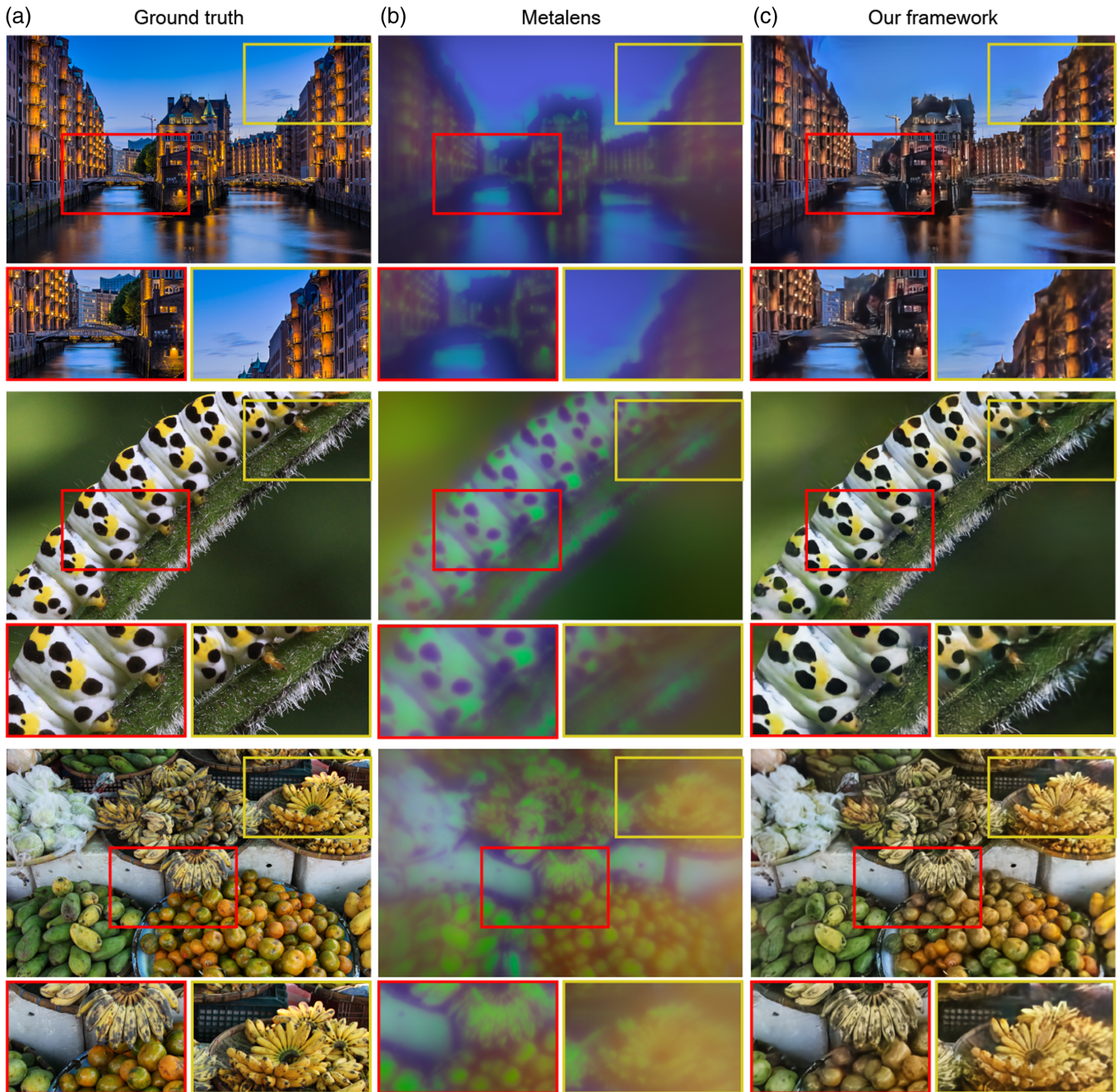
Contrastingly, the images reconstructed utilizing our proposed framework exhibit a remarkable fidelity to the ground truth across both peripheral and central regions, demonstrating the framework's proficiency in reinstating details obliterated by chromatic aberration. Such outcomes underscore the capability of our framework to surmount the intricate challenges posed by a highly irregular PSF, thereby significantly augmenting the imaging performance across a spectrum of scenarios. This denotes a substantial stride toward mitigating the complexities associated with aberration-induced degradation, heralding advancements in the fidelity and quality of imaging systems employing metalenses.

Despite the physical limitations inherent in metalenses, which cannot be overcome through conventional manufacturing processes alone, our application of deep learning enables imaging capabilities that exceed the physical performance limits of the metalenses. This innovative approach effectively bridges the gap between the inherent physical constraints and the desired imaging outcomes.

In the following sections, we present a comparative statistical analysis based on the test dataset to assess the quality of image restoration. This analysis further illustrates how our deep learning-enhanced framework not only compensates for the physical limitations of metalenses but also significantly improves the overall image quality.

#### 3.1 Quality of Image Restoration

Figure 5 comprehensively shows the results of the PSNR, structural similarity index measure (SSIM), learned perceptual image patch similarity (LPIPS) in RGB space, and mean absolute error (MAE) of the magnitudes, as well as cosine similarity (CS) in Fourier space calculated by comparing the metalens image and the image reconstructed by our framework with the ground truth image. The red horizontal lines in each box represent the median, and the boxes extend from the first to the third quartile. The whiskers span 1.5 times the interquartile range of the first and third quartiles. We conducted a statistical hypothesis test to ascertain whether the observed results exhibit statistically significant differences. This was accomplished through the utilization of a two-sided paired *t*-test to evaluate the performance disparity between images produced by metalenses and those reconstructed by the proposed framework. A significance level of  $P = 10^{-4}$  was set for the testing process.

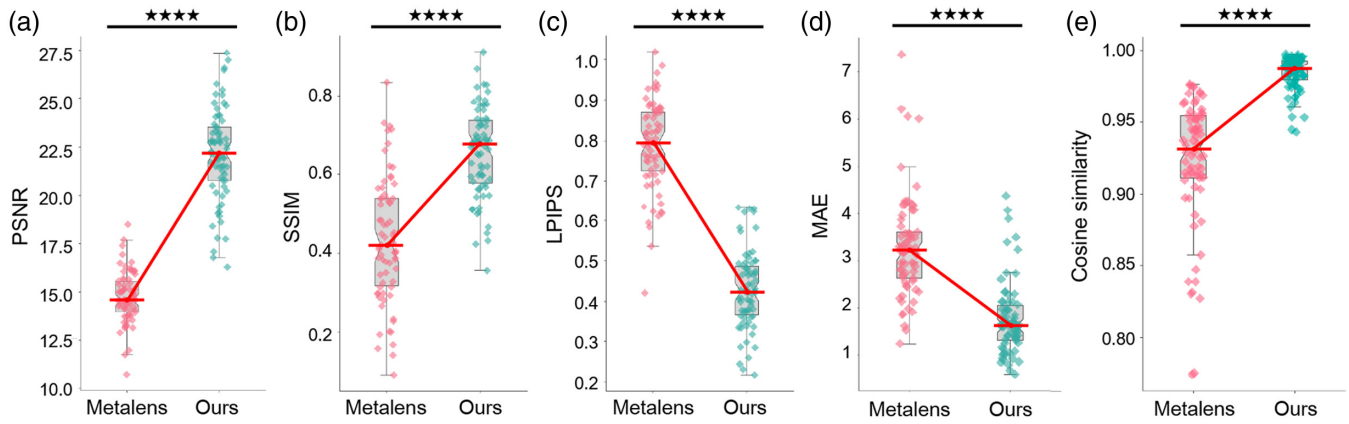


**Fig. 4** (a) Ground truth images, (b) metalens images, and (c) images reconstructed by our model. The images are affiliated with the test set data. The central (red) and outer (yellow) regions of the images are enlarged to access the restoration of the metalens image at high and low viewing angle, respectively. The outer regions of the metalens images (yellow box) are successfully restored, even though those are more severely degraded than the inner region (red box) due to the angular aberration under high viewing angle.

Within this analysis, the outcomes indicate a statistically significant variance across all evaluated metrics, as evidenced in Fig. 5. These metrics were assessed utilizing a test set comprising 70 data points. Also, Table 1 shows the quantitative results of the metalens imaging system, our framework, and state-of-the-art models for various metrics. The implications derived from each graph and the significance of the quantitative outcomes are elaborated below, providing a comprehensive analysis of the data and its relevance to the study's objectives.

To further understand the impact of our framework on the fidelity of image restoration, we examine PSNR and SSIM,<sup>59</sup> which serve as the foundational metrics. The former is a quantitative measure of the restoration quality of an image, calculated as the logarithmic ratio between the maximum possible power of a signal (image) and the power of corrupting noise that affects its fidelity. Higher PSNR values indicate better quality of the reconstructed image. The latter, SSIM, values the visual impact of three characteristics of an image: luminance, contrast,





**Fig. 5** Comparative statistical analysis of the proposed model and metalens imaging results using the test dataset. (a)–(e) Results of PSNR, SSIM, LPIPS in RGB space and CS, MAE of the magnitudes in Fourier space calculated by comparing the metalens image and the image reconstructed by our framework with the ground truth image. A statistical hypothesis test was performed through a two-sided paired  $t$ -test on the performance difference between the metalens image and the image reconstructed by our framework [significance level  $P = 10^{-4}$ , (a)  $1.055 \times 10^{-39}$ , (b)  $3.886 \times 10^{-35}$ , (c)  $1.363 \times 10^{-48}$ , (d)  $2.311 \times 10^{-35}$ , and (e)  $2.150 \times 10^{-38}$ ].

and structure, thus providing a more accurate reflection of perceived image quality.

Figure 5 presents a statistical analysis comparing the PSNR and SSIM values of the images captured through the metalens with those restored by our framework. As shown in Table 1, the framework showcased a remarkable improvement in image fidelity, elevating the PSNR by 7.37 dB and SSIM by 22.5%p compared to the original metalens images. These enhancements underscore our framework’s proficiency in mitigating the fidelity loss incurred by metalens aberrations, thus significantly elevating the quality of the reconstructed images closer to their ground truths.

While PSNR and SSIM are advantageous for assessing image fidelity and perceptual quality, they often fall short in evaluating the structured outputs. This limitation stems from their inability to fully capture the human visual system’s sensitivity to various image distortions, particularly in textured or detailed regions. To address this gap, LPIPS<sup>60</sup> was employed to evaluate the perceptual quality of the images. LPIPS evaluates perceptual similarity by utilizing pretrained deep learning networks (e.g., AlexNet), offering a nuanced measure that aligns more closely with human perception of image quality. Lower LPIPS values indicate better perceptual quality.

Table 1 demonstrates that our framework achieved a 35.6%p decrease in LPIPS, indicating a substantial enhancement in the perceptual resemblance of the reconstructed images to their original counterparts, as also observable in Fig. 5(c). This metric highlights the proposed framework’s capability to not only improve the objective quality of images but also their subjective, perceptual quality. We also compare our framework with state-of-the-art models, including restoration models for natural images (MIRNetv2,<sup>57</sup> HINet,<sup>58</sup> NAFNet<sup>38</sup>). As shown in Table 1, our framework surpasses these state-of-the-art models by a substantial margin in terms of PSNR, SSIM, and LPIPS. In addition, we conduct further experiments to measure and compare the restoration performance for spatially and spectrally varying degradations (Tables S4 and S5 in the [Supplementary Material](#)). This suggests that our framework is more suitable for the metalens image restoration task than conventional models designed for restoring natural images, such as those in the DIV2K dataset.<sup>51</sup>

The measured MTF of the metalens in Fig. 1(d) and qualitative results in Fig. 4(b) demonstrate intense degradation at high spatial frequencies. Consequently, it is crucial to restore the spatial-frequency information during the metalens image restoration task. It is pertinent to acknowledge that spatial

**Table 1** Comparison of quantitative assessments of various models using the test set of images ( $n = 70$ ). The first and second values of each column represent the mean and the standard deviation of the metrics, respectively. The best scores are marked as bold.

Model	Image quality metric			Assessment in frequency domain	
	PSNR	SSIM	LPIPS	MAE	CS
Metalens image	14.722/1.328	0.431/0.157	0.788/0.112	3.281/1.089	0.922/0.045
MIRNetv2	18.507/1.893	0.556/0.134	0.559/0.098	2.240/0.900	0.967/0.020
SFNet	18.223/1.727	0.567/0.129	0.519/0.095	2.194/0.837	0.965/0.020
HINet	21.364/2.333	0.641/0.121	0.456/0.097	1.851/0.800	0.982/0.013
NAFNet	21.689/2.382	0.642/0.120	0.440/0.097	1.817/0.801	0.983/0.013
Our framework	<b>22.095/2.423</b>	<b>0.656/0.114</b>	<b>0.432/0.096</b>	<b>1.759/0.779</b>	<b>0.984/0.012</b>

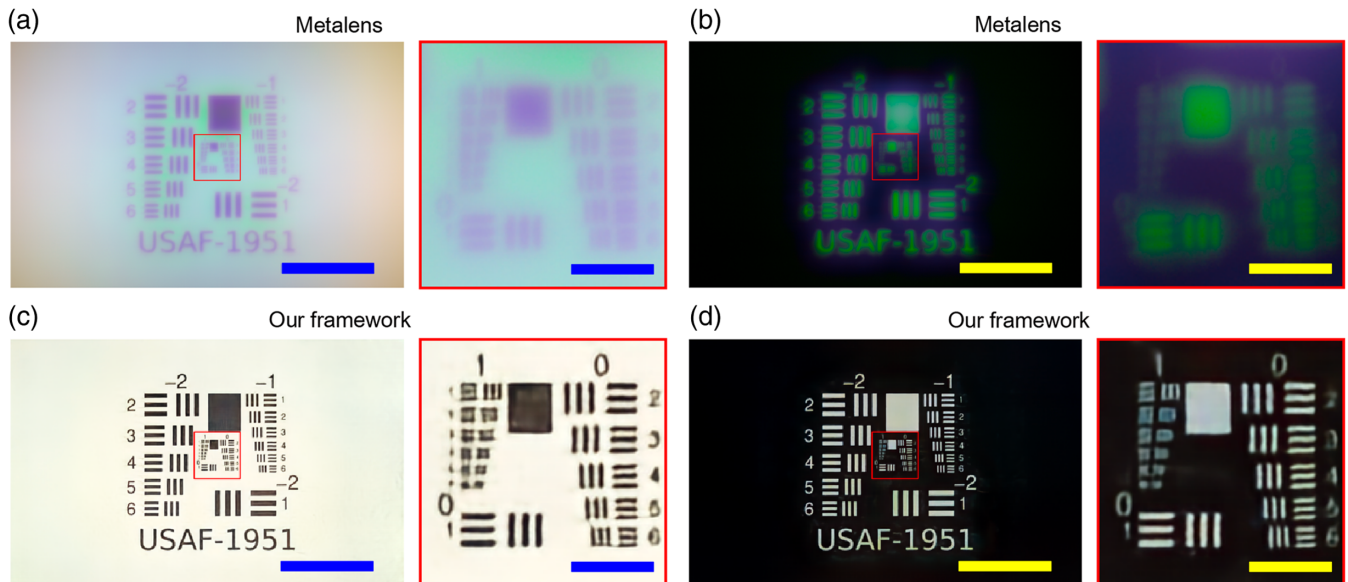
frequency can be represented as both magnitude and phase components, with the latter often heralded as important in signal processing realms.<sup>61</sup> We utilize two metrics to evaluate the magnitude and phase of the Fourier-transformed reconstructed images. In evaluating the fidelity of the reconstructed images, particularly concerning their frequency-dependent attributes, two metrics are employed: the MAE for assessing discrepancies in magnitude relative to the original images, and the CS for gauging phase congruence with the authentic images. These metrics are derived through the application of the FFT across images revitalized by disparate models. The ensuing MAE and CS metrics underscore a remarkable enhancement in image quality, as elucidated in Figs. 5(d) and 5(e) and Table 1. As shown in these figures, our framework demonstrates the superior performance of MAE and CS to the metalens imaging system and several state-of-the-art image restoration models in the frequency domain. Our framework achieved about twice the performance of the metalens imaging system for MAE and accomplished about 14%p for CS for the metalens images.

To demonstrate the restoration of the blur and color distortion visually, we tested our imaging system using 1951 U.S. Air Force resolution test chart images (USAF images). Figures 6(a) and 6(b) show monochromatic white and black USAF images captured by the metalens imaging system. These images exhibit severe blurring and strong color distortion, particularly showing greenish tints in white patterns. As shown in Figs. 6(c) and 6(d), the restored images illustrate that the pattern's colors are closer to white and black than the metalens images. Furthermore, the central regions of the images exhibit high sharpness, while the damaged images have severe blurring in these areas. Thus, our framework demonstrates superior prominence in enhancing overall image quality by achieving conspicuous color fidelity and sharpness.

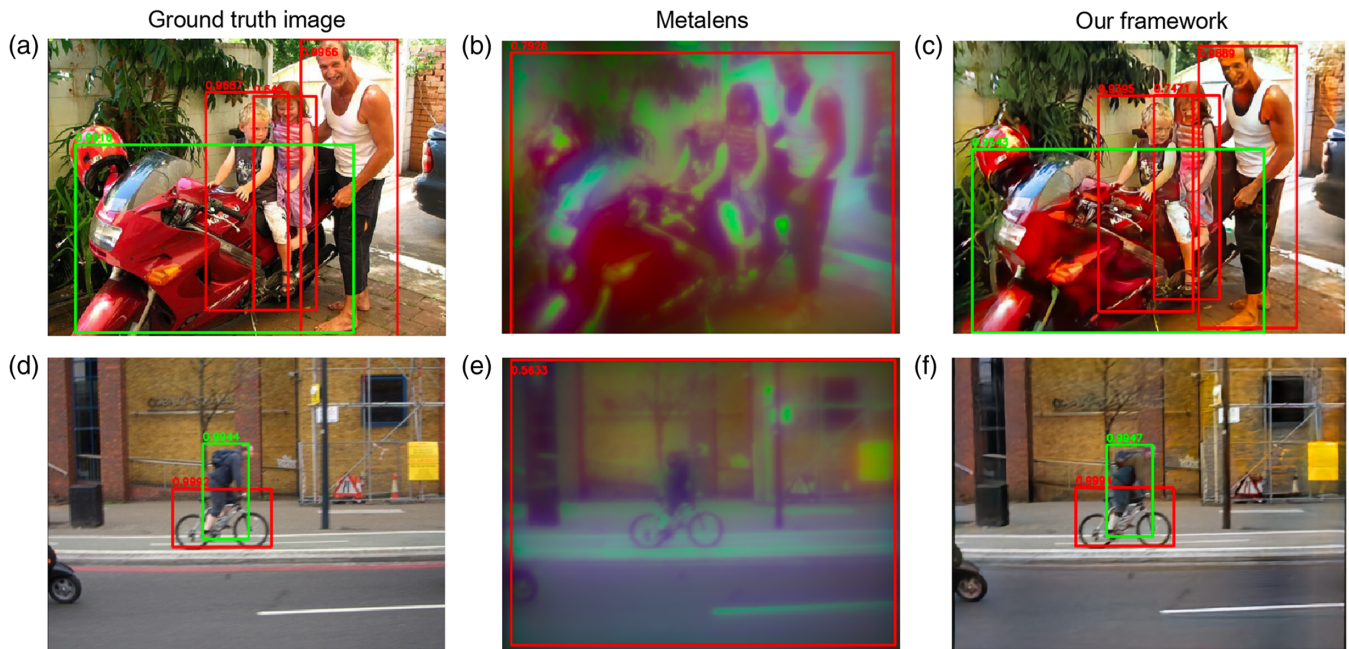
### 3.2 Object Detection Performance

We also assess the integrated system's utility beyond image quality enhancement by transitioning to one of the domains of practical applications, object detection. To validate the performance of our framework in object detection on the restored images, we first obtained a test dataset that consists of the ground truth, metalens, and restored images using the entire PASCAL VOC2007 dataset ( $n = 4952$ ).<sup>62</sup> This dataset comprises 4952 images with the information of the positions of objects and bounding box annotations for instances belonging to 20 different categories. Then, we employ a single shot multibox detector (SSD)<sup>63</sup> pre-trained for the PASCAL VOC2007 to detect bounding boxes of given images. Especially, we introduce mean average precision (AP) to evaluate object detection results. AP measures the model's accuracy in predicting the presence and correct localization of objects within an image. It provides a comprehensive assessment of the detection performance across varying thresholds of precision and recall, making it a standard benchmark in evaluating object detection algorithms. Specifically, AP is calculated in multiple scales using IoU thresholds ranging from 0.5 to 0.95. The terms "AP<sub>50</sub>" and "AP<sub>75</sub>" each represent the results of measurements with the IoU threshold set to 0.5 and 0.75, respectively.

Figure 7 shows examples of object detection using SSD. The detector predicts the entire region (red box) as an object because it cannot identify the detailed features in the metalens images [Fig. 6(b)]. On the other hand, the detector accurately predicts the bounding boxes at the desired objects in the restored images because compared to the original PASCAL VOC2007, the quality of the restored images is competitive [Figs. 6(a), 6(d) and 6(c), 6(f)]. Our AP, AP<sub>50</sub> on the restored images are ~34%p and 56%p higher than those on the metalens images. They are ~86% and 88% of the AP, AP<sub>50</sub> on the ground truth



**Fig. 6** (a) and (b) White and black USAF images captured by the metalens imaging system, respectively. (c) and (d) White and black USAF images restored by our framework, respectively. The image in the red boxes shows the enlarged image in the central region indicated as red box. The scale bars in the original and enlarged images are 3 and 0.5 mm, respectively, indicating the distance on the image sensor.



**Fig. 7** Object detection results using a pre-trained SSD model on (a), (d) the original images, (b), (e) the metalens images, and (c), (f) the images restored by our framework. The pre-trained SSD model could not detect any objects in the metalens images accurately; however, it successfully captured multiple classes and objects in images restored by our framework.

(Table S2 in the [Supplementary Material](#)). The improvement in AP scores for our framework-restored images compared to the original metalens images signifies the restoration's practical implications. Higher AP scores on our framework-restored images indicate that the model effectively recovers enough detail and structure from the aberrated metalens images to facilitate accurate object detection, closely approximating the performance on ground truth images. This enhancement is particularly crucial for applications in autonomous navigation, surveillance, and augmented reality, where precise object detection is paramount.

## 4 Conclusion

In this study, we have demonstrated DNN-based image restoration framework for large-area mass-produced metalenses. Our approach effectively mitigates the severe chromatic and angular aberrations inherent in large-area broadband metalenses, a challenge that has long impeded the widespread adoption of the metalenses. Also, assuming the uniform quality of mass-produced metalenses, the optimized restoration model can be applied to other metalenses manufactured at the same process. By employing an adversarial learning scheme in the Fourier space coupled with positional embedding, we have transcended traditional limitations, enabling the restoration of high-spatial-frequency information and facilitating aberration-free, full-color imaging through mass-produced metalenses. The profound implications of our findings extend a commercially viable pathway toward the development of ultra-compact, efficient, and aberration-free imaging systems.

## Disclosures

The authors declare that there are no financial interests, commercial affiliations, or other potential conflicts of interest that could have influenced the objectivity of this research or the writing of this paper.

## Code and Data Availability

The metalens and ground truth image dataset can be accessed from the GitHub repository at: [https://github.com/yhy258/EIDL\\_DRMI](https://github.com/yhy258/EIDL_DRMI) and Figshare repository at: <https://doi.org/10.6084/m9.figshare.24634740.v1>. USAF images taken with the metalens, reconstructed with the NAFNet and reconstructed with our framework, are also available in the GitHub repository. The test set of PASCAL VOC2007 dataset can be found at: <http://host.robots.ox.ac.uk/pascal/VOC/voc2007/>.

The code in this study is available in the GitHub repository at: [https://github.com/yhy258/EIDL\\_DRMI](https://github.com/yhy258/EIDL_DRMI). The pre-trained models are also available in the same repository. We used the publicly available SSD code (<https://github.com/amdegroot/ssd.pytorch>) for object detection using the PyTorch library.

## Author Contributions

H. C. and J. R. supervised and designed the study. J. S., J. J., C. K., and J. H. developed the methodology for the idea and performed the experiments. J. J. and J. S. acquired the data. J. Kim, J. S., J. J., S. M., E. L., and J. Kang analyzed the metalens and image data. J. Kim and J. R. designed and fabricated the metalenses. H. C., J. R., J. S., J. J., J. Kim, and J. Kang wrote the paper.

## Acknowledgments

This work was supported by a grant from the National Research Foundation of Korea (NRF), funded by the Korean government (MSIT) (RS-2024-00338048) and also supported by the Culture, Sports and Tourism R&D Program through a grant from the Korea Creative Content Agency funded by the Ministry of Culture, Sports and Tourism in 2024 (RS-2024-00332210), as

well as the Artificial Intelligence Graduate School Program (RS-2020-II201373, Hanyang University) supervised by the IITP, and under the Artificial Intelligence Semiconductor Support Program to nurture the best talents (IITP(2024)-RS-2023-00253914), funded by the Korea government. J.R. acknowledges the POSCO-POSTECH-RIST Convergence Research Center program funded by POSCO, the Samsung Research Funding & Incubation Center for Future Technology grant (SRFC-IT1901-52) funded by Samsung Electronics, the POSTECH-Samsung Semiconductor Research Center (IO201215-08187-01) funded by Samsung Electronics, the NRF grants (RS-2024-00462912, RS-2024-00416272, RS-2024-00337012, RS-2024-00408446, NRF-2022M3H4A1A02074314) funded by the MSIT of the Korean government, and the Korea Evaluation Institute of Industrial Technology (KEIT) grant (1415185027/20019169, Alchemist project) funded by the Ministry of Trade, Industry and Energy (MOTIE) of the Korean government. J. Kim acknowledges the Asan Foundation Biomedical Science fellowship, and the Presidential Science fellowship by the MSIT of the Korean government. E.L. acknowledges the SBS Foundation fellowship, the Presidential Science fellowship funded by the MSIT of the Korean government, and the NRF M.S. fellowship (RS-2024-00464712) funded by the Ministry of Education of the Korean government.

## References

1. S. Wang et al., "A broadband achromatic metalens in the visible," *Nat. Nanotechnol.* **13**(3), 227–232 (2018).
2. W. T. Chen et al., "A broadband achromatic metalens for focusing and imaging in the visible," *Nat. Nanotechnol.* **13**(3), 220–226 (2018).
3. F. Yang et al., "Wide field-of-view metalens: a tutorial," *Adv. Photonics* **5**, 033001 (2023).
4. C.-Y. Fan, C.-P. Lin, and G.-D. J. Su, "Ultrawide-angle and high-efficiency metalens in hexagonal arrangement," *Sci. Rep.* **10**(1), 15677 (2020).
5. M. Khorasaninejad et al., "Metalenses at visible wavelengths: diffraction-limited focusing and subwavelength resolution imaging," *Science* **352**(6290), 1190–1194 (2016).
6. S. Shrestha et al., "Broadband achromatic dielectric metalenses," *Light: Sci. Appl.* **7**(1), 85 (2018).
7. J. Kim et al., "One-step printable platform for high-efficiency metasurfaces down to the deep-ultraviolet region," *Light: Sci. Appl.* **12**(1), 68 (2023).
8. J. Kim et al., "Metasurface holography reaching the highest efficiency limit in the visible via one-step nanoparticle-embedded-resin printing," *Laser Photon. Rev.* **16**(8), 2200098 (2022).
9. J. Kim et al., "A water-soluble label for food products prevents packaging waste and counterfeiting," *Nat. Food* **5**(4), 293–300 (2024).
10. J. Kim et al., "Dynamic hyperspectral holography enabled by inverse-designed metasurfaces with oblique helicoidal cholesterics," *Adv. Mater.* **36**(24), 2311785 (2024).
11. J. Kim et al., "8" wafer-scale, centimeter-sized, high-efficiency metalenses in the ultraviolet," *Mater. Today* **73**, 9–15 (2024).
12. S.-W. Moon et al., "Wafer-scale manufacturing of near-infrared metalenses," *Laser Photon. Rev.* **18**, 2400929 (2024).
13. J. Kim et al., "Scalable manufacturing of high-index atomic layer-polymer hybrid metasurfaces for metaphotonics in the visible," *Nat. Mater.* **22**(4), 474–481 (2023).
14. Y. Zhou et al., "Flat optics for image differentiation," *Nat. Photonics* **14**(5), 316–323 (2020).
15. M. Y. Shalaginov et al., "Single-element diffraction-limited fish-eye metalens," *Nano Lett.* **20**(10), 7429–7437 (2020).
16. A. Martins et al., "On metalenses with arbitrarily wide field of view," *ACS Photonics* **7**(8), 2073–2079 (2020).
17. F. Wang et al., "Visible achromatic metalens design based on artificial neural network," *Adv. Opt. Mater.* **10**(3), 2101842 (2022).
18. M. K. Chen et al., "Meta-lens in the sky," *IEEE Access* **10**, 46552–46557 (2022).
19. S.-J. Kim et al., "Dielectric metalens: properties and three-dimensional imaging applications," *Sensors* **21**(13), 4584 (2021).
20. E. Choi et al., "360° structured light with learned metasurfaces," *Nat. Photonics* **18**(8), 848–855 (2024).
21. Z. Li et al., "Meta-optics achieves RGB-achromatic focusing for virtual reality," *Sci. Adv.* **7**, eabe4458 (2021).
22. Z. Li et al., "Inverse design enables large-scale high-performance meta-optics reshaping virtual reality," *Nat. Commun.* **13**(1), 2409 (2022).
23. H. Liang et al., "Ultrahigh numerical aperture metalens at visible wavelengths," *Nano Lett.* **18**(7), 4460–4466 (2018).
24. O. D. Miller and Z. Kuang, "Fundamental limits for large-area meta-optics," in *Freeform Optics*, p. JTu6A.6 (2021).
25. S. M. Kamali et al., "A review of dielectric optical metasurfaces for wavefront control," *Nanophotonics* **7**(6), 1041–1068 (2018).
26. J. Engelberg and U. Levy, "Optimizing the spectral range of diffractive metalenses for polychromatic imaging applications," *Opt. Express* **25**(18), 21637–21651 (2017).
27. F. Presutti and F. Monticone, "Focusing on bandwidth: achromatic metalens limits," *Optica* **7**(6), 624–631 (2020).
28. S. So et al., "Revisiting the design strategies for metasurfaces: fundamental physics, optimization, and beyond," *Adv. Mater.* **35**, e2206399 (2022).
29. D. Zhang et al., "Cascaded chiral birefringent media enabled planar lens with programmable chromatic aberration," *Photonix* **5**, 17 (2024).
30. E. Tseng et al., "Neural nano-optics for high-quality thin lens imaging," *Nat. Commun.* **12**, 6493 (2021).
31. R. Maman et al., "Achromatic imaging systems with flat lenses enabled by deep learning," *ACS Photonics* **10**(12), 4494–4500 (2023).
32. Y. Dong et al., "Achromatic single metalens imaging via deep neural network," *ACS Photonics* **11**(4), 1645–1656 (2024).
33. J. E. Fröch et al., "Beating bandwidth limits for large aperture broadband nano-optics," arXiv:2402.06824 (2024).
34. H. Chung and O. D. Miller, "High-NA achromatic metalenses by inverse design," *Opt. Express* **28**(5), 6945–6965 (2020).
35. Z. Lin and S. G. Johnson, "Overlapping domains for topology optimization of large-area metasurfaces," *Opt. Express* **27**(22), 32445–32453 (2019).
36. L. Huang et al., "Design and analysis of extended depth of focus metalenses for achromatic computational imaging," *Photonics Res.* **8**(10), 1613–1623 (2020).
37. M. Khorasaninejad et al., "Achromatic metalens over 60 nm bandwidth in the visible and metalens with reverse chromatic dispersion," *Nano Lett.* **17**(3), 1819–1824 (2017).
38. L. Chen et al., "Simple baselines for image restoration," in *Eur. Conf. Comput. Vision*, pp. 17–33 (2022).
39. S. W. Zamir et al., "Restormer: efficient transformer for high-resolution image restoration," in *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit.*, pp. 5728–5739 (2022).
40. G. Yoon et al., "Printable nanocomposite metalens for high-contrast near-infrared imaging," *ACS Nano* **15**(1), 698–706 (2021).
41. J. W. Goodman, *Introduction to Fourier Optics*, Roberts and Company Publishers (2005).
42. Y. Zhou et al., "Image restoration for under-display camera," in *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit.*, pp. 9179–9188 (2021).
43. F. Shen and A. Wang, "Fast-Fourier-transform based numerical integration method for the Rayleigh-Sommerfeld diffraction formula," *Appl. Opt.* **45**(6), 1102–1110 (2006).
44. Y. Li et al., "NTIRE 2023 challenge on image denoising: methods and results," in *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit.*, pp. 1904–1920 (2023).

45. S. Nah et al., “NTIRE 2021 challenge on image deblurring,” in *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit.*, pp. 149–165 (2021).
46. J. S. Goldstein, I. S. Reed, and L. L. Scharf, “A multistage representation of the Wiener filter based on orthogonal projections,” *IEEE Trans. Inf. Theory* **44**(7), 2943–2959 (1998).
47. M. T. Heath, *Scientific Computing: An Introductory Survey*, SIAM (2018).
48. R. Yang et al., “NTIRE 2022 challenge on super-resolution and quality enhancement of compressed video: dataset, methods and results,” in *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit.*, pp. 1221–1238 (2022).
49. W. Wu et al., “URetInex-Net: Retinex-based deep unfolding network for low-light image enhancement,” in *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit.*, pp. 5901–5910 (2022).
50. Q. Jiang et al., “Annular computational imaging: capture clear panoramic images through simple lens,” *IEEE Trans. Comput. Imaging* **8**, 1250–1264 (2022).
51. E. Agustsson and R. Timofte, “NTIRE 2017 Challenge on single image super-resolution: dataset and study,” in *IEEE Conf. Comput. Vision Pattern Recognit. Workshops*, pp. 1122–1131 (2017).
52. S. Nah, T. Hyun Kim, and K. Mu Lee, “Deep multi-scale convolutional neural network for dynamic scene deblurring,” in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, pp. 3883–3891 (2017).
53. X. Chu et al., “Improving image restoration by revisiting global information aggregation,” in *Eur. Conf. Comput. Vision*, pp. 53–71 (2022).
54. L. Ruthotto and E. Haber, “An introduction to deep generative modeling,” *GAMM-Mitteilungen* **44**(2), e202100008 (2021).
55. T. Miyato et al., “Spectral normalization for generative adversarial networks,” arXiv:1802.05957 (2018).
56. J. H. Lim and J. C. Ye, “Geometric GAN,” arXiv:1705.02894 (2017).
57. S. W. Zamir et al., “Learning enriched features for fast image restoration and enhancement,” *IEEE Trans. Pattern Anal. Mach. Intell.* **45**(2), 1934–1948 (2022).
58. L. Chen et al., “HINet: half instance normalization network for image restoration,” in *Proc. IEEE/CVF Conf. Comput. Vision Pattern Recognit.*, pp. 182–192 (2021).
59. Z. Wang et al., “Image quality assessment: from error visibility to structural similarity,” *IEEE Trans. Image Process.* **13**(4), 600–612 (2004).
60. R. Zhang et al., “The unreasonable effectiveness of deep features as a perceptual metric,” in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, pp. 586–595 (2018).
61. A. V. Oppenheim and J. S. Lim, “The importance of phase in signals,” *Proc. IEEE* **69**, 529–541 (1981).
62. M. Everingham et al., “The PASCAL visual object classes (VOC) challenge,” *Int. J. Comput. Vis.* **88**(2), 303–338 (2010).
63. W. Liu et al., “SSD: single shot multibox detector,” in *Eur. Conf. Comput. Vision*, pp. 21–37 (2016).
64. Y. Cui et al., “Selective frequency network for image restoration,” presented at *Eleventh Int. Conf. Learn. Represent.* (2022).
65. Y. Zhang et al., “Image super-resolution using very deep residual channel attention networks,” in *Proc. Eur. Conf. Comput. Vision (ECCV)*, pp. 286–301 (2018).

**Joonhyuk Seo** has been a graduate student at Hanyang University since 2023, specializing in deep learning. He received his BS degree in media technology from Hanyang University in 2023. He is conducting research applying AI to optics, focusing on developing strong image restoration methods and novel electromagnetic surrogate solvers for fast simulation.

**Jaegang Jo** is a PhD student in the Electronic Engineering Department at Hanyang University. He received his BS degree from the Physics Department at Sungkyunkwan University and his MS degree at the Graduate School of Convergence Science and Technology from Seoul National University.

**Jooheon Kim** received his BS degree in mechanical engineering in 2021, and then started his integrated MS/PhD program in mechanical engineering at Pohang University of Science and Technology (POSTECH). His research is mainly focused on metasurface, its nanofabrication, and its practical applications.

**Joonho Kang** is a PhD student in the Department of Artificial Intelligence Semiconductor Engineering at Hanyang University. He received his BS degree in electronic engineering from Hanyang University in 2024. His research focuses on inverse design and optimization of metamaterials and nanophotonic devices, as well as advanced simulations in nanophotonics to enhance design efficiency and device performance.

**Chanik Kang** is a PhD student in the Department of Artificial Intelligence, Hanyang University, Republic of Korea. He received his BS degree in mechanical engineering from Soongsil University in 2022. His current research focuses on deep-learning photonics and surrogate simulations for large-area photonic design.

**Seong-Won Moon** is a PhD candidate in mechanical engineering at POSTECH since 2020. He received his MS degree in 2020 and his BS degree in 2018 in electrical engineering from Kyungpook National University. His research interests are metasurfaces, metalenses, holograms, and the applications of metasurfaces, including novel imaging systems.

**Eunji Lee** received her BS degree in 2023 in chemical engineering at POSTECH, Republic of Korea. She is currently an MS and PhD integrated candidate under the guidance of Prof. Junsuk Rho at POSTECH. Her research interests include the metasurfaces and their application, nanofabrication, and space photonics.

**Je Hyeong Hong** received his BA and MEng in electrical and information sciences and his PhD in information engineering from the University of Cambridge, United Kingdom, where he collaborated with Microsoft and Toshiba. He completed postdoctoral research at KIST in Republic of Korea during alternative military service in 2018 to 2021. Since 2021, he has been an assistant professor in electronic engineering at Hanyang University. His research interests include computer vision, machine learning, and optimization.

**Junsuk Rho** is currently a Mu-Eun-Jae endowed chair professor with a joint appointment in mechanical engineering, chemical engineering, and electrical engineering at POSTECH. His research is focused on developing novel nanophotonic materials and devices based on fundamental physics and experimental studies of deep sub-wavelength light-matter interaction. He received his BS (2007), MS (2008), and PhD (2013) degrees all in mechanical engineering at Seoul Nation University, University of Illinois, Urbana-Champaign, and University of California, Berkeley, respectively.

**Haejun Chung** has been an assistant professor at Hanyang University since 2022, specializing in inverse design for photonic applications. He received his BS degree in electrical engineering from Illinois Institute of Technology in 2010 and his MS (2013) and PhD (2017) degrees from Purdue University. He conducted postdoctoral research at Yale University, developing fast inverse design algorithms and metasurfaces, and later at MIT, focusing on large-area metalenses and tunable metasurfaces.