



OPEN Metalens-style image synthesis for metalens imaging via image-to-image translation

Chanik Kang^{1,6}, Hyewon Suk^{2,6}, Joonhyuk Seo³, Ikbeom Jang⁴✉ & Haejun Chung^{1,2,5}✉

Metalenses offer wafer-scale, ultra-thin optics for compact cameras, but strong chromatic and field-dependent aberrations still limit their practical use. Deep learning-based aberration correction can restore high-quality images from metalens captures, but current pipelines typically require hundreds to thousands of paired images per device. We address this data bottleneck by formulating metalens aberration synthesis as a deterministic, metalens-conditioned image-to-image translation problem. A generator is trained on a dataset of paired metalens and conventional images from a mass-producible metalens, then used to transform photographs into metalens-style outputs that reproduce realistic chromatic aberration, field-dependent blur, and spatial distortion. On a test set, the proposed translator reduces LPIPS(VGG) from 0.305 to 0.117 (~62%) compared with a state-of-the-art transformer-based restoration baseline. Once trained, the translator can generate 600 synthetic metalens-style images in roughly 30 s on a single GPU, versus about 30 min for real metalens acquisition, a ~60× reduction in data-collection time. These synthetic pairs alone suffice to train a metalens image restoration model, suggesting that our approach can help alleviate the data bottleneck in future metalens imaging research.

Keywords Computational imaging, Metalens, Synthesis image, Image-to-image translation, Data augmentation

Metalenses are planar metasurfaces composed of dense arrays of subwavelength scatterers that tailor the wavefront by imparting spatially varying phase, amplitude, and polarization^{1–3}. In contrast, as illustrated in Fig. 1a, conventional refractive singlets and compound lens stacks rely on millimeter–centimeter-scale thickness and multiple glass elements to achieve a given numerical aperture and to correct aberrations^{3,4}. At high numerical apertures (NA), these refractive objectives often become bulky and inefficient⁵. They require thick, heavy glass stacks and carefully optimized multi-element modules to keep images sharp and to avoid characteristic forms of blur and distortion^{6,7}.

By implementing a customized wavefront within a subwavelength-scale thickness, metalenses offer the potential to replace or complement such bulky optics with compact, wafer-scale elements that support high-NA and wide field of view^{8–10}. Successful demonstrations include broadband or achromatic metalenses, multifunctional metasurface cameras, and compact modules for microscopy and wearable or mobile imaging^{11–18}.

Instead of bending light through bulk refraction in thick glass elements, metalenses implement the required phase profile by designing meta-atoms to satisfy a desired wavefront. This meta-atom-based phase control, realized through subwavelength scatterers, enables high NA in a thin, lithography-compatible form factor and relaxes packaging constraints compared with multi-element refractive stacks^{19,20}. However, individual meta-atoms exhibit strongly wavelength-, incidence-angle-, and polarization-dependent responses^{21–23}. In particular, the point-spread function (PSF), which captures the impulse response of the lens, can vary significantly with wavelength, field angle, and polarization, leading to pronounced chromatic dispersion and rapidly changing image quality across the field of view^{24,25}. As a result, metalenses are sensitive to fabrication tolerances, alignment errors, and deviations from the design spectrum²⁶. For general-purpose imaging, these effects manifest as design-

¹Department of Artificial Intelligence, Hanyang University, Seoul 04763, Republic of Korea. ²Department of Artificial Intelligence Semiconductor Engineering, Hanyang University, Seoul 04763, Republic of Korea. ³Department of Electrical Engineering and Computer Science, University of California Irvine, Irvine 92697, USA. ⁴Department of Computer Engineering, Hankuk University of Foreign Studies, Yongin 17035, Republic of Korea. ⁵Department of Electronic Engineering, Hanyang University, Seoul 04763, Republic of Korea. ⁶Chanik Kang and Hyewon Suk have contributed equally to this work. ✉email: ijang@hufs.ac.kr; haejun@hanyang.ac.kr

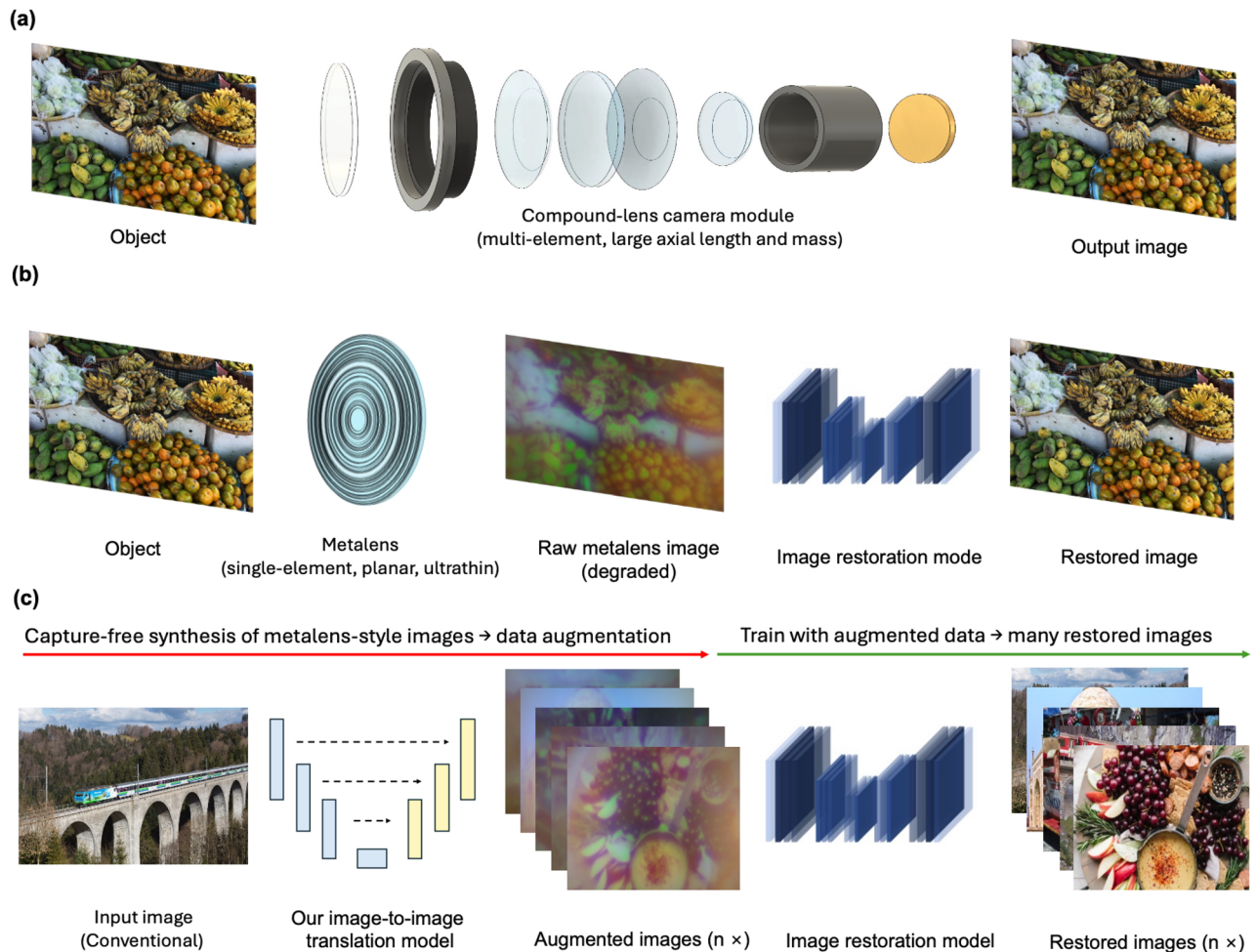


Fig. 1. Conventional, metalens, and proposed augmentation pipelines. **(a)** Conventional imaging system: A compound refractive module with multiple lens elements produces high-quality images at the cost of increased axial length and mass. **(b)** Ultra-compact metalens imaging system: A single metalens yields raw captures with chromatic fringes, field-dependent blur, and spatial distortion; a learned restoration network is required to obtain a conventional-looking image. **(c)** Proposed augmentation pipeline: A PSF-free, device-conditioned image-to-image translator converts ordinary photographs into metalens-style renders that reproduce chromatic and field-dependent artifacts. The resulting synthetic pairs enable training of restoration models without additional metalens data collection, providing many restored images for downstream analysis.

specific aberrations, wavelength-dependent focal shifts, field-dependent blur, edge darkening (vignetting), and spatially varying distortion that degrade image quality and hinder feature detection, recognition, and tracking²⁷.

Despite impressive progress in metalens design, several challenges remain before metalenses can serve as drop-in replacements for conventional camera modules. On the optics side, dispersion-engineered and achromatic metalenses can partially correct chromatic focal shifts^{11,13}. However, they typically trade off bandwidth^{21,28}, NA²³, and field of view^{6,29}, or require complex multi-layer²³ or multi-surface architectures that are difficult to fabricate at scale³⁰. Hybrid systems that combine metalenses with refractive corrector elements can improve aberration performance, but they sacrifice some of the thickness and weight advantages that motivate metasurface optics in the first place³¹. Moreover, metalens performance is often tightly coupled to a specific wavelength band, sensor, and packaging configuration, which limits portability across applications and product generations³².

To address these limitations, there has been growing interest in computational metalens imaging that jointly considers optics and post-processing^{33–36}. As illustrated in Fig. 1b, physics-based pipelines estimate or measure the PSF of a given metalens and then apply deconvolution or model-based restoration to reduce blur and color fringing^{24,37}. However, for metalenses intended for commercial, high-volume production, software correction of metalens-specific aberrations still requires large image datasets captured with the manufactured metalens modules²⁴. Building such per-device datasets across wafers, production batches, and alignment conditions is a labor-intensive, time-consuming process that is difficult to integrate directly into standard high-throughput manufacturing flows.

In this work, we target this data bottleneck from a complementary direction. Instead of repeatedly collecting massive paired datasets for every new metalens design, we develop a few-shot data-augmentation framework that learns to synthesize *metalens-style* images, defined as ordinary photographs transformed to mimic the characteristic aberrations and degradations of a given metalens. Given a clean input image captured by a conventional lens, the proposed image-to-image translation model generates a plausible counterpart that exhibits the chromatic aberration, field-dependent blur, and spatial distortion produced by a target metalens. Trained on a modest calibration set of metalens/conventional pairs, the model then scales to produce large volumes of synthetic training data without explicit PSF measurement, wave-propagation simulation, or exhaustive metalens image acquisition (Fig. 1c). In this way, our approach combines the advantages of deep learning-based image restoration with the compact, wafer-thin form factor of metalenses, making software correction of metalens aberrations more compatible with high-volume manufacturing.

Results

We now evaluate how well the proposed image-to-image translator can synthesize realistic *metalens-style* degradations and how useful the resulting synthetic images are for metalens image restoration. Here, we use *metalens-style* to denote ordinary photographs transformed to mimic the characteristic chromatic fringes, field-dependent blur, and spatial distortions produced by a target metalens.

From a computer-vision perspective, our task lies between data augmentation and image-to-image translation. Classical augmentation pipelines (random crops, flips, rotations, color jitter, and policy-based schemes) are designed to increase the stochastic diversity of training sets^{38,39}, while restoration and super-resolution networks are trained to map low-quality inputs to high-quality outputs^{40,41}. In meta-optical imaging, recent deep-learning-based systems have mainly followed this enhancement direction, restoring strongly aberrated metalens captures to near-conventional-lens quality using hundreds to thousands of paired images per device^{24,26}. In contrast to these enhancement models, we focus on the reverse direction, mapping clean images to degraded *metalens-style* images. We deliberately degrade conventional images to resemble the outputs of a specific, mass-producible metalens, aiming to generate large synthetic datasets that can replace or significantly reduce the need for labor-intensive paired collections.

We organize our evaluation into four parts. First, we quantify the quality of the proposed *metalens-style* synthesis on a test set of paired conventional and metalens images using standardized image metrics. Second, we benchmark the translator against conventional restoration and super-resolution baselines trained under identical preprocessing and evaluation protocols to verify that a degradation-focused model is necessary to reproduce metalens-specific artifacts. Throughout this section, the restoration/super-resolution baselines are trained and evaluated in the clean \rightarrow metalens direction, and their performance is interpreted in terms of degradation-synthesis fidelity to the ground-truth metalens images rather than enhancement quality. Third, we perform ablation studies on the loss formulation and upsampling strategy to isolate their effects on chromatic aberration reproduction, field-dependent blur, and texture retention. Finally, we assess the utility of the synthesized datasets for metalens image restoration by training a standard restorer exclusively on translated images and measuring its generalization to real metalens captures. This last experiment directly tests our central hypothesis that physically informed, metalens-conditioned augmentation can alleviate the data bottleneck in deep-learning-driven meta-optical imaging.

Image synthesis for metalens imaging

We first evaluate the fidelity of the proposed *metalens-style* synthesis on a set of paired conventional and metalens-captured images. As detailed in Fig. 2, our translator is trained on a small calibration set of paired captures and then applied to unseen conventional photographs to generate *metalens-style* renderings. As illustrated in Fig. 2, the generator employs a U-Net architecture⁴², leveraging skip connections to effectively capture detailed information between the input and output images, thereby generating accurately transformed images. While the original Pix2Pix⁴³ model typically employs an L1 loss function (mean absolute error, MAE)⁴⁴, this study utilizes an L2 loss function (mean squared error, MSE)⁴⁵ to preserve realistic image textures better. The Discriminator employs a robust 1x1 PatchGAN architecture⁴³, which enhances its ability to discern detailed textures and assess the realism of generated images with respect to colorfulness.

Once trained, the translator can convert arbitrary conventional photographs into *metalens-style* images that exhibit realistic optical aberrations, including chromatic fringes and field-dependent blur. In our implementation, generating a dataset of 600 synthetic *metalens-style* images takes on the order of 30 s, compared with roughly 30 min for real metalens acquisition (Table 1).

To quantify the fidelity of our synthetic metalens augmentation, we compare three classes of methods—image-to-image translation, image restoration, and super-resolution—on a test dataset of paired conventional and metalens-captured images. Each model is evaluated using peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM), and learned perceptual image patch similarity (LPIPS). As shown in Table 2, Pix2Pix outperforms all competitors, achieving a mean PSNR of 29.37 dB, a mean SSIM of 0.9877, an LPIPS score of 0.0477, and an LPIPS-VGG score of 0.1168. Qualitative results in Fig. 3 corroborate the quantitative findings: Pix2Pix faithfully reproduces chromatic fringes, field-dependent blur, and spatially varying distortions, closely matching ground-truth metalens captures without introducing perceptual artifacts. In comparison, restoration-based methods tend to over-sharpen or smooth critical features, and super-resolution models like LAPAR⁴⁶ are unable to emulate realistic aberrations.

Restoration using synthetic training data

We next investigate whether realistic restoration of metalens photographs can be achieved using only a limited amount of real metalens data together with the proposed augmentation. We first train the translator on about

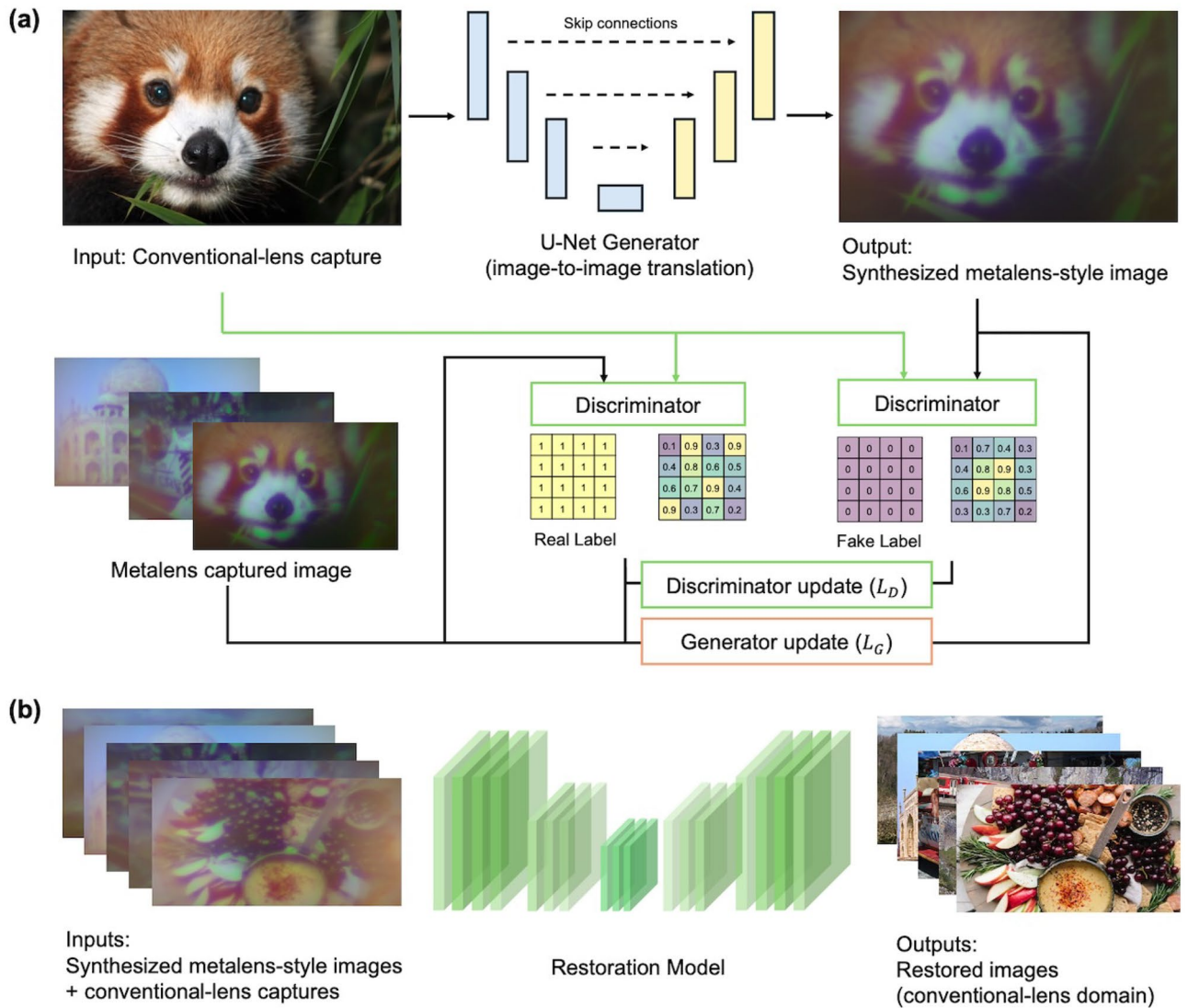


Fig. 2. Data synthesis and restoration workflow for metalens imaging. **(a)** Pix2Pix-based conditional translation framework. A U-Net generator learns to convert conventional-lens captures into synthesized metalens-style images that exhibit realistic chromatic and field-dependent aberrations. A pair of discriminators evaluates realism at multiple patch scales, with the generator and discriminators updated alternately for stable training. **(b)** Restoration pipeline using synthetic data. The restoration network is trained exclusively on paired data composed of synthesized metalens-style images and their corresponding conventional-lens references, producing restored outputs in the conventional-lens domain without real metalens supervision.

Source	# images	Time per image [s]	Total time [s]
Real metalens capture	600	3.0	1800 (\approx 30 min)
Synthetic generation (ours)	600	0.05	30 (\approx 0.5 min)

Table 1. Acquisition time for real metalens datasets versus synthetic metalens-style datasets for $N = 600$ images. The synthetic generation time is assumed to be ~ 50 ms per image on a single NVIDIA RTX 4090 GPU.

600 real metalens images. Then, images from the Real-SR dataset⁵² are cropped to a resolution of 512×360 and converted into metalens-style images, yielding 500 synthetic samples. The restoration network is trained on these synthetic data, along with 1,100 real metalens images. Training pairs consist of either a real metalens image or a metalens-style synthetic image paired with its corresponding clean reference image. Restoration performance is evaluated on real metalens photographs that are not used during training.

Model	SSIM \uparrow	PSNR \uparrow	LPIPS(VGG) \downarrow
Image-to-Image Translation			
Ours	0.988/0.007	29.373/3.022	0.117/0.050
Palette ⁴⁷	0.968/0.016	14.322/3.948	0.600/0.196
Image Restoration			
Restormer ⁴¹	0.969/0.0150	30.653/2.987	0.305/0.051
Fourier domain model ⁴⁸	0.787/0.102	22.624/2.583	0.335/0.061
NAFNET ⁴⁹	0.929/0.035	25.984/2.042	0.436/0.045
Super Resolution			
LAPAR ⁴⁶	0.876/0.038	17.898/2.607	0.600/0.048
Generative model			
Autoencoder ⁵⁰	0.829/0.062	16.611/1.854	0.488/0.058
Variational Autoencoder ⁵¹	0.814/0.095	14.322/3.948	0.439/0.158

Table 2. Performance comparison of image models on metalens data, grouped by task. For each metric, the first value denotes the mean and the second value denotes the standard deviation; higher SSIM and PSNR are better, whereas lower LPIPS is better.

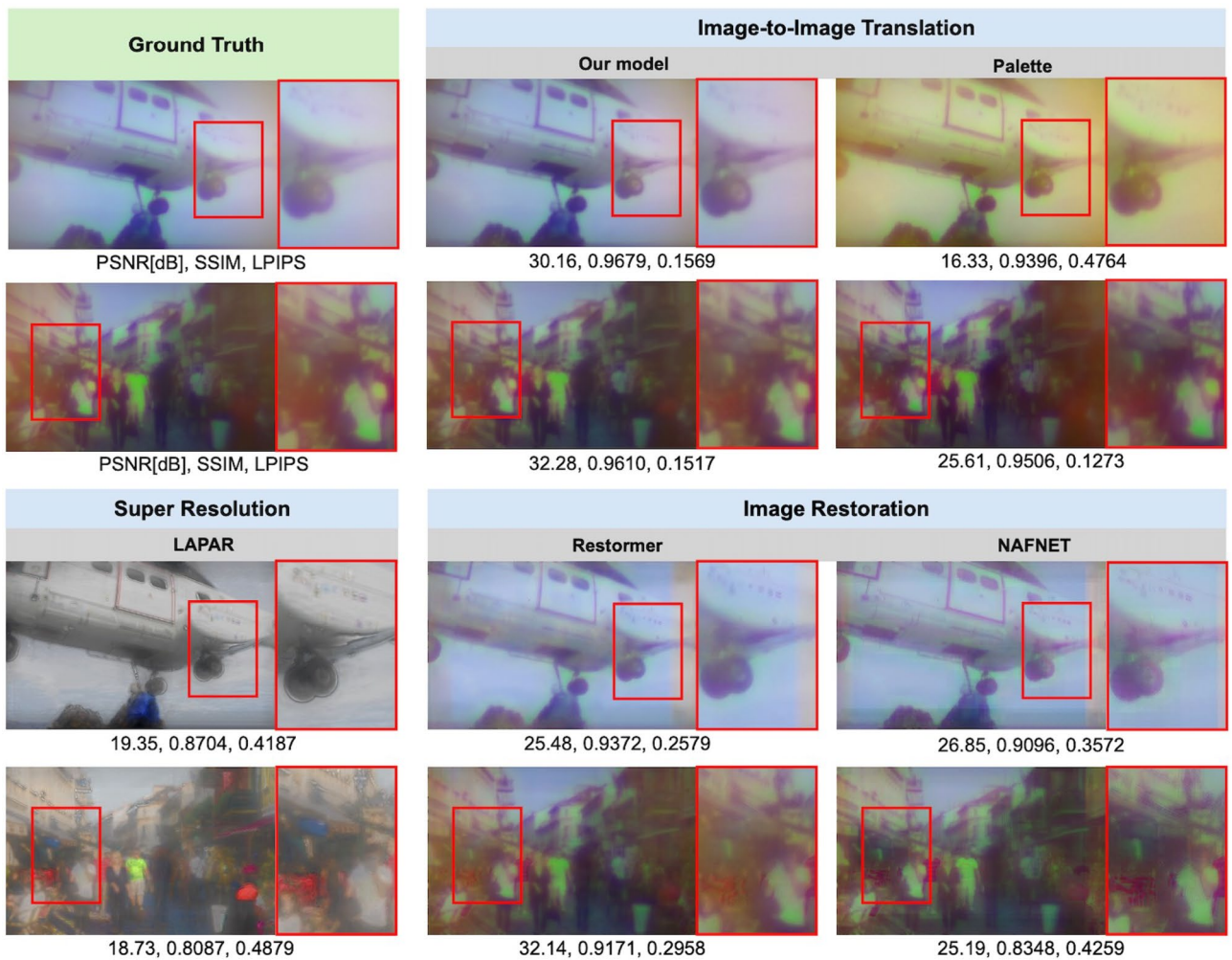


Fig. 3. Comparison of metalens-style synthesis across models. The proposed Pix2Pix-based translator most faithfully reproduces chromatic aberration and field-dependent blur without introducing noticeable perceptual artifacts, yielding more vivid color and realistic contrast than restoration-based or super-resolution methods, which often produce desaturated and over-smoothed regions.

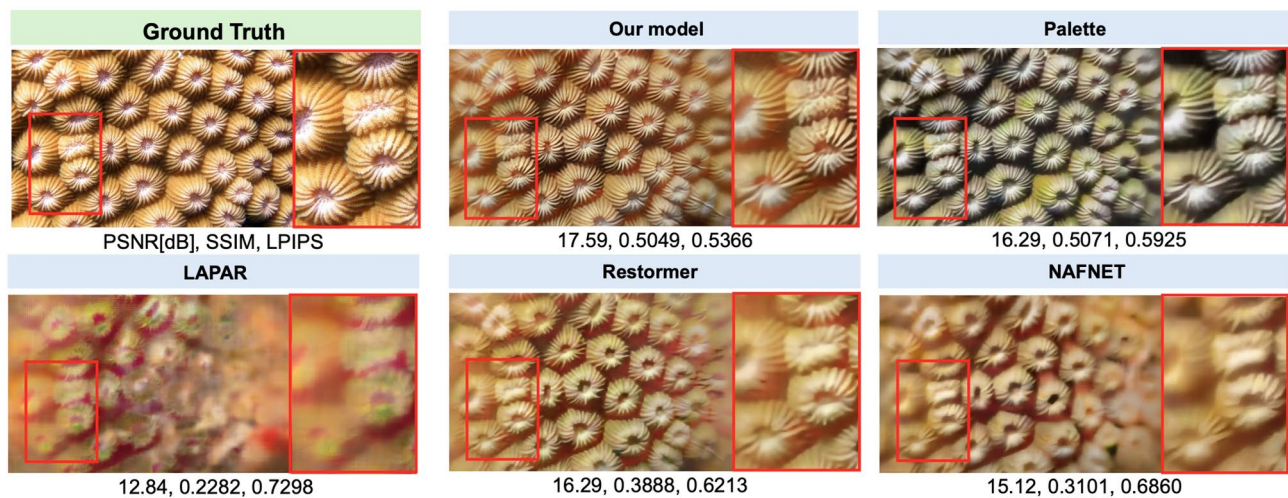


Fig. 4. Qualitative restoration results on real metalens photos using a model trained on real and augmented data. Right: restoration results for real inputs. Major structures and color tones are recovered even with only limited real-pair supervision; finer textures remain challenging, especially near the image edges.

Model	PSNR \uparrow	SSIM \uparrow	LPIPS(VGG) \downarrow
Ours	19.159	0.745	0.452
Palette ⁴⁷	17.185	0.732	0.654
Restormer ⁴¹	17.640	0.698	0.543
NAFNET ⁴⁹	15.474	0.601	0.654
LAPAR ⁴⁶	11.829	0.481	0.766

Table 3. Restoration performance on the real metalens + synthetic metalens test set. Higher values indicate better performance for PSNR and SSIM, while lower values are better for LPIPS.

Evaluation on real metalens images

Generalization was assessed on a collection of photographs captured with the metalens that were not used in translator training or restoration training. Although the restorer is trained only on synthetic pairs generated from a limited amount of real measurements, the augmentation-only restorer consistently recovers the main structures and overall color balance in real metalens images. As shown in the visual comparison in Fig. 4, the Pix2Pix-based model produces more natural, structurally faithful outputs than the other baselines, with better fine-detail preservation and perceptual quality (SSIM, LPIPS). While reconstructions tend to be somewhat softer than those obtained by a restorer trained directly on real pairs, the results demonstrate that synthetic augmentation alone can still provide meaningful restoration for real metalens inputs.

Quantitative results

For quantitative validation, we evaluated restoration models on 70 real metalens test pairs using PSNR, SSIM, and LPIPS metrics. All restorers were trained exclusively on synthetic pairs composed of translated metalens-style inputs and their corresponding conventional-lens references; no real metalens/conventional pairs were used during training.

As summarized in Table 3, the restorer achieves a PSNR of 19.159, SSIM of 0.745, and LPIPS of 0.452. This outperforms strong restoration baselines trained under the same synthetic-only regimen: the next-best PSNR (17.640 dB) and LPIPS (0.543) are obtained by Restormer⁴¹, while the next-best SSIM (0.732) is achieved by Palette⁴⁷. In other words, our model improves PSNR by roughly 1.5 dB over the strongest baseline and reduces LPIPS by about 0.09 in absolute terms, indicating better perceptual agreement with conventional-lens references even though it is trained solely on synthetic metalens-style pairs without access to any real metalens/conventional pairs.

These quantitative gains are consistent with the qualitative results in Fig. 4: the our Pix2Pix-based restorer recovers global structure and color tone more faithfully than competing methods and suppresses many of the severe chromatic aberration and field-dependent blur present in the raw metalens captures. Taken together, these results support the use of synthesized metalens-style data as effective training targets for restoration in metalens imaging.

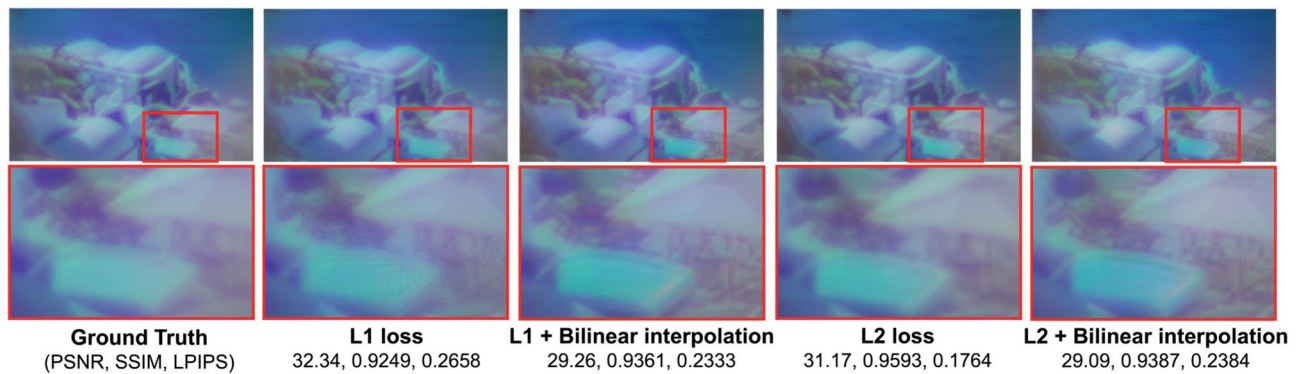


Fig. 5. Effect of loss and upsampling on metalens-style synthesis. We compare L1 versus L2 objectives and transpose-convolution versus bilinear upsampling; the L2 + bilinear variant best suppresses checkerboard artifacts while preserving subtle aberration features.

Number of data	PSNR↑	SSIM↑	LPIPS↓
100	20.623	0.884	0.322
300	21.905	0.921	0.228
500	22.196	0.934	0.196
600	29.373	0.988	0.117

Table 4. Effect of training set size on metalens-style synthesis performance. Results are reported for PSNR, SSIM, and LPIPS on the test set; higher PSNR and SSIM and lower LPIPS indicate better performance.

Observed failure modes and limitations

A consistent artifact appears near the image periphery. At large field angles, edges and fine detail can be over-blurred in the restored outputs. This behavior is attributed to underrepresented position-dependent blur in the augmented training inputs. Because the current augmentation is not explicitly conditioned on spatial coordinates, the restorer tends to learn a spatially uniform correction that does not capture edge-of-field behavior²⁴. In practice, boundaries in the synthetic images are not blurred as strongly or as anisotropically as those observed in real captures, and the restoration model reproduces this discrepancy. Remedies include injecting normalized spatial coordinates into both translator and restorer, sampling training patches by field angle, and employing spatially varying losses. Mixing a small number of real pairs during restorer training is also expected to reduce any residual gap between synthetic and real data while keeping the collection cost low.

Ablation study

Effect of loss and upsampling

We isolate the effects of loss formulation and upsampling approach through an ablation study illustrated in Fig. 5. Specifically, we train the Pix2Pix⁴³ model with either an L1 or L2 objective and replace transposed convolutions with bilinear upsampling. Quantitative and qualitative assessments reveal that the L2 objective paired with bilinear interpolation reduces checkerboard artifacts and preserves subtle aberration features more faithfully than other combinations. In contrast, models trained with an L1 objective or using transpose convolution introduce visible grid patterns or over-smooth fine details, underscoring the importance of loss choice and sampling method for realistic metalens augmentation.

Effect of training set size

We examined data efficiency by training the translator with 100, 300, 500, and 600 paired samples, as summarized in Table 4. As the set size increases, PSNR and SSIM improve while LPIPS decreases, indicating progressively better agreement with real metalens captures. With smaller training sets (100–300 pairs), the generator largely preserves hue and overall style but struggles near the image edges: straight lines bend or break, thin contours fray, and high-frequency textures collapse into overly smooth bands. As the training pool grows, these artifacts are progressively reduced: boundary warping recedes, line continuity improves, and small features such as wires and edge highlights are retained more reliably. At the largest set size (600 pairs), the synthesized images exhibit more stable geometry and sharper micro-details across scenes, indicating that additional data primarily benefits spatial fidelity rather than color rendition and reduces rare failure cases at object edges and near field boundaries.

Discussion

We formulate deterministic aberration synthesis for metalens imaging as an image-to-image translation problem. The learned translator maps clean images to metalens-style renderings that reproduce chromatic

fringing, field-dependent blur, and spatial distortions, and the resulting synthetic data enable training and evaluation of restoration models on real captures. Rather than enumerating additional metrics, this section interprets the implications, situates the method relative to prior work, and outlines directions for improvement. With respect to prior work in image-to-image translation, most methods emphasize stochastic diversity and multi-modality^{53–55}. By contrast, our objective is metalens-conditioned fidelity, that is, faithfully reproducing the aberrations of a specific metalens from a clean input image.

Generality across metalens designs and operating conditions. In this study, we train and evaluate the translator using data captured with a single metalens design (consistent with the lens used in our prior metalens image restoration setting) to maintain a controlled and directly comparable experimental setup. Because metalens imaging characteristics can vary across designs, fabrication tolerances, and system-level alignment, the current validation does not establish zero-shot generalization to arbitrary metalenses. In practical deployment, we expect a modest re-calibration step (e.g., fine-tuning with a small paired conventional/metalens calibration set) when the metalens design or fabrication/alignment regime changes. As follow-up work, we aim to extend the evaluation to multiple metalens designs and diverse conditions to systematically assess robustness and generality.

The proposed formulation reduces or even eliminates the need for dedicated physical calibration experiments, while remaining compatible with physics-based pipelines. While classical PSF modeling or rigorous simulation provides interpretability, such approaches can be costly or brittle for metasurfaces whose PSFs vary rapidly with field angle and wavelength and depend on nanoscale geometry.

Our approach offers three practical benefits. First, it lowers the entry cost for new devices by amortizing limited calibration into a generator that synthesizes large labeled sets without explicit PSF estimation. Second, it preserves subtle, spatially varying artifacts that generic style-transfer objectives tend to suppress, thereby improving the ecological validity of downstream training. Third, it provides a practical framework for optics–algorithm co-design: after a modest per-device calibration, scene-matched synthetic datasets enable controlled, low-cost iteration over metalens form factors and post-capture processing.

Taken together, these steps aim to sharpen peripheral detail, stabilize color under severe fringing, and position PSF-free aberration synthesis as a practical complement to physics-based modeling for metalens imaging. Beyond the laboratory, such a framework could help compress multi-element refractive modules into single-metalens front-ends with learned digital correction, reducing lens count, thickness, and assembly complexity in mobile and embedded cameras⁵⁶. By providing metalens-specific synthetic datasets on demand, it can shorten the iteration loop between optical design^{57,58}, wafer-level manufacturing⁵⁹, and on-device imaging⁷, thereby supporting co-designed hardware–software imaging systems in which flat optics and neural networks jointly deliver the performance traditionally achieved by bulky glass stacks.

Methods

Evaluation protocol

To evaluate the proposed augmentation framework, we compare the images generated by our Pix2Pix⁴³-based translator with outputs from an image-to-image translation baseline⁴⁷, two state-of-the-art restoration models^{41,49}, and leading super-resolution models⁴⁶. Although restoration and super-resolution networks are originally designed for enhancement (low-quality → high-quality), we include them here as enhancement-oriented controls and retrain/use them in the clean → metalens synthesis direction under identical paired-data supervision and preprocessing. This benchmark therefore evaluates how well each model can reproduce metalens-specific degradations; accordingly, the metrics in Table 2 quantify synthesis fidelity to the ground-truth metalens captures rather than enhancement quality.

For quantitative assessment, we compute PSNR, SSIM, and LPIPS between each model's synthetic outputs and the corresponding ground-truth metalens captures. Table 2 reports these metrics on a test set, and Fig. 3 presents representative qualitative examples. To verify that our synthetic images serve as valid training data, we train a standard restoration network using only the generated dataset and then evaluate its performance on real metalens images. We measure PSNR, SSIM, and LPIPS for these reconstructions and analyze their error distributions. The resulting scores demonstrate that models trained on our augmented data achieve competitive reconstruction quality, confirming that the generated images are effective for downstream learning in metalens imaging tasks.

Dataset detail

In our framework, we use two types of image data: (i) ground-truth images captured with a conventional lens, and (ii) their corresponding captures obtained with a metalens. We employ the same dataset as in our previous work by Seo et al., “Deep-learning-driven end-to-end metalens imaging”²⁴, which comprises 600 training pairs and 70 test pairs.

Hyperparameters

Training hyperparameters followed settings in the Table 5.

Training pairs	600
Patch/crop size	128
Batch size	16
Epochs	7500
Learning rate/schedule	3e-4/CosineAnnealingLR ⁶⁰
Optimizer	AdamW ⁶¹
Loss	PSNR Loss

Table 5. Training setup for the Pix2Pix-based metalens-style translator.

Data availability

The metalens and ground truth image dataset can be accessed from the Figshare repository at: <https://doi.org/10.6084/m9.figshare.24634740.v1>.

Received: 20 November 2025; Accepted: 9 January 2026

Published online: 20 January 2026

References

- Khorasaninejad, M. & Capasso, F. Metalenses: Versatile multifunctional photonic components. *Science* **358**, eaam8100 (2017).
- Barulin, A. et al. Dual-wavelength metalens enables epi-fluorescence detection from single molecules. *Nature Commun.* **15**, 26 (2024).
- Shrestha, S., Overvig, A. C., Lu, M., Stein, A. & Yu, N. Broadband achromatic dielectric metalenses. *Light: Sci. Appl.* **7**, 85 (2018).
- Martins, A. et al. Correction of aberrations via polarization in single layer metalenses. *Adv. Opt. Mater.* **10**, 2102555 (2022).
- Lalanne, P. & Chavel, P. Metalenses at visible wavelengths: Past, present, perspectives. *Laser Photo. Rev.* **11**, 1600295 (2017).
- Shastri, K. & Monticone, F. Bandwidth bounds for wide-field-of-view dispersion-engineered achromatic metalenses. *EPJ Appl. Metamater.* **9** (2022).
- Zeng, Y., Zhong, H., Long, Z., Cao, H. & Jin, X. From performance to structure: A comprehensive survey of advanced metasurface design for next-generation imaging. *npj Nanophotonics* **2**, 39 (2025).
- Khorasaninejad, M. et al. Metalenses at visible wavelengths: Diffraction-limited focusing and subwavelength resolution imaging. *Science* **352**, 1190–1194 (2016).
- Arbabi, A. & Faraon, A. Advances in optical metalenses. *Nature Photonics* **17**, 16–25 (2023).
- Choi, M. et al. Roll-to-plate printable rgb achromatic metalens for wide-field-of-view holographic near-eye displays. *Nature Materials* 1–9 (2025).
- Chen, W. T. et al. A broadband achromatic metalens for focusing and imaging in the visible. *Nature Nanotechnol.* **13**, 220–226 (2018).
- Lee, G.-Y. et al. Metasurface eyepiece for augmented reality. *Nature Commun.* **9**, 4562 (2018).
- Wang, S. et al. A broadband achromatic metalens in the visible. *Nature Nanotechnol.* **13**, 227–232 (2018).
- Arbabi, E. et al. Two-photon microscopy with a double-wavelength metasurface objective lens. *Nano Lett.* **18**, 4943–4948 (2018).
- Chen, S., Liu, W., Li, Z., Cheng, H. & Tian, J. Metasurface-empowered optical multiplexing and multifunction. *Adv. Mater.* **32**, 1805912 (2020).
- Li, Y. et al. Ultracompact multifunctional metalens visor for augmented reality displays. *Photonix* **3**, 29 (2022).
- Brongersma, M. L. et al. The second optical metasurface revolution: moving from science to technology. *Nature Rev. Electr. Eng.* **2**, 125–143 (2025).
- Chung, H., Zhang, F., Li, H., Miller, O. D. & Smith, H. I. Inverse design of high-na metalens for maskless lithography. *Nanophotonics* **12**, 2371–2381 (2023).
- Pan, M. et al. Dielectric metalens for miniaturized imaging systems: Progress and challenges. *Light: Sci. Appl.* **11**, 195 (2022).
- Yun, J.-G. et al. Compact eye camera with two-third wavelength phase-delay metalens. *Nature Commun.* **16**, 7299 (2025).
- Presutti, F. & Monticone, F. Focusing on bandwidth: Achromatic metalens limits. *Optica* **7**, 624–631 (2020).
- Mansouree, M. et al. Multifunctional 2.5 d metastructures enabled by adjoint optimization. *Optica* **7**, 77–84 (2020).
- Chung, H. & Miller, O. D. High-na achromatic metalenses by inverse design. *Optics Express* **28**, 6945–6965 (2020).
- Seo, J. et al. Deep-learning-driven end-to-end metalens imaging. *Adv. Photonics* **6**, 066002–066002 (2024).
- Yang, F. et al. Wide field-of-view metalens: a tutorial. *Adv. Photonics* **5**, 033001–033001 (2023).
- Tseng, E. et al. Neural nano-optics for high-quality thin lens imaging. *Nature Commun.* **12**, 6493 (2021).
- Zou, X. et al. Imaging based on metalenses. *Photonix* **1**, 2 (2020).
- Chen, Q., Gao, Y., Pian, S. & Ma, Y. Theory and fundamental limit of quasiachromatic metalens by phase delay extension. *Phys. Rev. Lett.* **131**, 193801 (2023).
- Engelberg, J., Mazurski, R. & Levy, U. Nature inspired design methodology for a wide field of view achromatic metalens. *Nanophotonics* (2025).
- Liang, H. et al. High performance metalenses: Numerical aperture, aberrations, chromaticity, and trade-offs. *Optica* **6**, 1461–1470 (2019).
- Shih, K.-H. & Renshaw, C. K. Hybrid meta/refractive lens design with an inverse design using physical optics. *Appl. Optics* **63**, 4032–4043 (2024).
- Yang, Y. et al. Nanofabrication for nanophotonics. *ACS nano* **19**, 12491–12605 (2025).
- Lin, Z. et al. End-to-end metasurface inverse design for single-shot multi-channel imaging. *Optics Express* **30**, 28358–28370 (2022).
- Wei, K. et al. Large-area fabrication-aware computational diffractive optics. arXiv preprint [arXiv:2505.22313](https://arxiv.org/abs/2505.22313) (2025).
- Dong, Y. et al. Full-color, wide field-of-view metalens imaging via deep learning. *Adv. Optical Mater.* **13**, 2402207 (2025).
- Kim, Y. & Kim, I. Broadband achromatic metalens for high-resolution imaging. *Light: Sci. Appl.* **14**, 204 (2025).
- Park, Y. et al. End-to-end optimization of metalens for broadband and wide-angle imaging. *Adv. Optical Mater.* **13**, 2402853 (2025).
- Shorten, C. & Khoshgofaar, T. M. A survey on image data augmentation for deep learning. *J. Big Data* **6**, 1–48 (2019).
- Cubuk, E. D., Zoph, B., Mane, D., Vasudevan, V. & Le, Q. V. Autoaugment: Learning augmentation policies from data. arXiv preprint [arXiv:1805.09501](https://arxiv.org/abs/1805.09501) (2018).
- Dong, C., Loy, C. C., He, K. & Tang, X. Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **38**, 295–307 (2015).

41. Zamir, S. W. *et al.* Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5728–5739 (2022).
42. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 234–241 (Springer, 2015).
43. Isola, P., Zhu, J.-Y., Zhou, T. & Efros, A. A. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1125–1134 (2017).
44. Zhao, H., Gallo, O., Frosio, I. & Kautz, J. Loss functions for image restoration with neural networks. *IEEE Transactions on Computational Imaging* **3**, 47–57 (2016).
45. Dong, C., Loy, C. C., He, K. & Tang, X. Learning a deep convolutional network for image super-resolution. In *European Conference on Computer Vision*, 184–199 (Springer, 2014).
46. Li, W. *et al.* Lapar: Linearly-assembled pixel-adaptive regression network for single image super-resolution and beyond. *Adv. Neural Inf. Process. Syst.* **33**, 20343–20355 (2020).
47. Saharia, C. *et al.* Palette: Image-to-image diffusion models. In *ACM SIGGRAPH 2022 Conference Proceedings*, 1–10 (2022).
48. Jiang, L., Dai, B., Wu, W. & Loy, C. C. Focal frequency loss for image reconstruction and synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 13919–13929 (2021).
49. Chen, L., Chu, X., Zhang, X. & Sun, J. Simple baselines for image restoration. In *European Conference on Computer Vision*, 17–33 (Springer, 2022).
50. Hinton, G. E. & Salakhutdinov, R. R. Reducing the dimensionality of data with neural networks. *Science* **313**, 504–507 (2006).
51. Kingma, D. P. & Welling, M. Auto-encoding variational bayes. arXiv preprint [arXiv:1312.6114](https://arxiv.org/abs/1312.6114) (2013).
52. Cai, J., Zeng, H., Yong, H., Cao, Z. & Zhang, L. Toward real-world single image super-resolution: A new benchmark and a new model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 3086–3095 (2019).
53. Richardson, E. *et al.* Encoding in style: A stylegan encoder for image-to-image translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2287–2296 (2021).
54. Gatys, L. A., Ecker, A. S. & Bethge, M. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2414–2423 (2016).
55. Murez, Z., Kolouri, S., Kriegman, D., Ramamoorthi, R. & Kim, K. Image to image translation for domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4500–4509 (2018).
56. Hou, M., Chen, Y., Li, J. & Yi, F. Single 5-centimeter-aperture metalens enabled intelligent lightweight mid-infrared thermographic camera. *Sci. Adv.* **10**, eado4847 (2024).
57. Kang, C., Seo, D., Boriskina, S. V. & Chung, H. Adjoint method in machine learning: A pathway to efficient inverse design of photonic devices. *Mater. Des.* **239**, 112737 (2024).
58. Kang, C. *et al.* Large-scale photonic inverse design: Computational challenges and breakthroughs. *Nanophotonics* **13**, 3765–3792 (2024).
59. Park, J.-S. *et al.* All-glass 100 mm diameter visible metalens for imaging the cosmos. *ACS Nano* **18**, 3187–3198 (2024).
60. Loshchilov, I. & Hutter, F. Sgdr: Stochastic gradient descent with warm restarts. arXiv preprint [arXiv:1608.03983](https://arxiv.org/abs/1608.03983) (2016).
61. Loshchilov, I. & Hutter, F. Decoupled weight decay regularization. arXiv preprint [arXiv:1711.05101](https://arxiv.org/abs/1711.05101) (2017).

Author contributions

This project was initiated by C.K. and H.S., Experiment design was carried out by C.K. and H.S., supported by J.S., Model training was performed by H.S.. This project was supervised by I.J. and H.C.. The manuscript was prepared by C.K., H.S. and H.C., C.K., H.S., I.J. and H.C. reviewed the manuscript.

Funding

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (RS-2024-00338048, RS-2024-00414119, RS-2025-25463760); by the Ministry of Science and ICT (MSIT) of the Republic of Korea under the Global Research Support Program in the Digital Field (RS-2024-00412644) supervised by the Institute of Information and Communications Technology Planning & Evaluation (IITP); by the Culture, Sports and Tourism R&D Program through the Korea Creative Content Agency grant funded by the Ministry of Culture, Sports and Tourism in 2024 (RS-2024-00332210); by the Artificial Intelligence Graduate School Program (No. RS-2020-II201373, Hanyang University) supervised by IITP; by the artificial intelligence semiconductor support program to nurture the best talents (IITP-(2025)-RS-2023-00253914) funded by the Korean government; and by MSIT (RS-2025-02218723, RS-2025-02283217).

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to I.J. or H.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026