RESEARCH ARTICLE

ADVANCED OPTICAL MATERIALS
www.advopticalmat.de

Full-Color, Wide Field-of-View Metalens Imaging via Deep Learning

Yunxi Dong, Bowen Zheng, Fan Yang, Hong Tang, Huan Zhao, Yi Huang, Tian Gu, Juejun Hu, and Hualiang Zhang**

Chromatic aberration has been the main showstopper for metalenses when it comes to imaging applications with broadband sources such as ambient light. In wide field-of-view metalenses, this challenge becomes far more severe due to exacerbated lateral chromatic aberrations. In this paper, it is demonstrated, for the first time, full-color wide field-of-view imaging using a fisheye metalens coupled with deep learning computational processing. This approach is capable of restoring panoramic images with enhanced signal-to-noise ratio while effectively correcting chromatic aberration, distortion, and vignetting. Furthermore, it is shown that the deep learning algorithm is robust against various lighting conditions and object distances, making it a versatile solution for practical imaging applications involving wide field-of-view metalenses.

1. Introduction

Using nanostructures much smaller than the wavelength to control light propagation, metalenses represent a transformative advance in optics technology. Unlike traditional lenses which rely on the bulk properties of materials to bend light, metalenses utilize arrays of nanostructures to achieve fine control over the phase, amplitude, and polarization of incoming light waves.[1,2] Metalenses have been shown to be highly effective in correcting monochromatic aberrations such as spherical aberration, astigmatism, coma, and field curvature.[3,4] In particular, several recent studies have identified ways to suppress these aberrations across a large field-of-view (FOV) to achieve diffraction-limited performance.[5–11]

Nonetheless, chromatic aberration, which fundamentally arises from diffraction by the Fresnel zones,[12] remains the Achilles heel for metalenses. Significant efforts have been devoted to realizing achromatic metalenses (all with a restricted FOV) via dispersion engineering or zone engineering,[13–15] although these approaches are bound by the accessible time delays.[16,17] Moreover, aberration correction becomes even more challenging for wide-FOV (WFOV) metalenses. Unlike narrow-FOV optics where only longitudinal chromatic aberration (i.e., wavelength-dependent change of focal length) needs to be compensated, the additional lateral chromatic aberration (i.e., wavelength-dependent change of image height) imposes a far more severe adverse impact on image quality in WFOV systems.[18–20] These challenges have thus far precluded experimental demonstration of broadband imaging using WFOV metalenses.

To ameliorate this inherent shortcoming of metalenses, image post-processing has been shown to be an effective means to expand their spectral bandwidth.[21] Deep learning (DL) models have also been successfully applied to realizing achromatic flat lens imaging using the UNet-based single achromatic metalens model,[22] the transformer-based flat lens model,[23] the generative adversarial network (GAN)-based circularly polarized metalens model.[24] and the transformer-neural network approach for WFOV imaging model.[25] However, these prior studies are constrained by certain limitations, such as the need for predefined point spread function (PSF) datasets and limited FOV. The efficacy of image post-processing algorithms remains an open question in the presence of large lateral chromatic aberration.

Our work here aims to fill in the gap by combining deep-learning-based post-processing with a WFOV fisheye metalens to demonstrate, for the first time, full-color WFOV imaging using a metalens. In our study, the Comprehensive Image Transformation Model (CITM) [22] was employed as a foundational framework to analyze the preprocessing requirements and address nonlinearities inherent in image restoration. After data collection, we implement the MIRUNet [22] model to directly correct chromatic aberration and other image degradation mechanisms in WFOV metalens system. Leveraging a unique convolutional neural network optimized for image enhancement, the model is trained on a comprehensive WFOV dataset to ensure broad

Y. Dong, B. Zheng, H. Tang, H. Zhao, Y. Huang, H. Zhang
Department of Electrical and Computer Engineering
University of Massachusetts Lowell
Lowell, MA 01854, USA
E-mail: hujuejun@mit.edu
F. Yang, T. Gu, J. Hu
Department of Materials Science and Engineering
Massachusetts Institute of Technology
Cambridge, MA 02139, USA
E-mail: hualiang_zhang@uml.edu

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/adom.202402207

DOI: 10.1002/adom.202402207

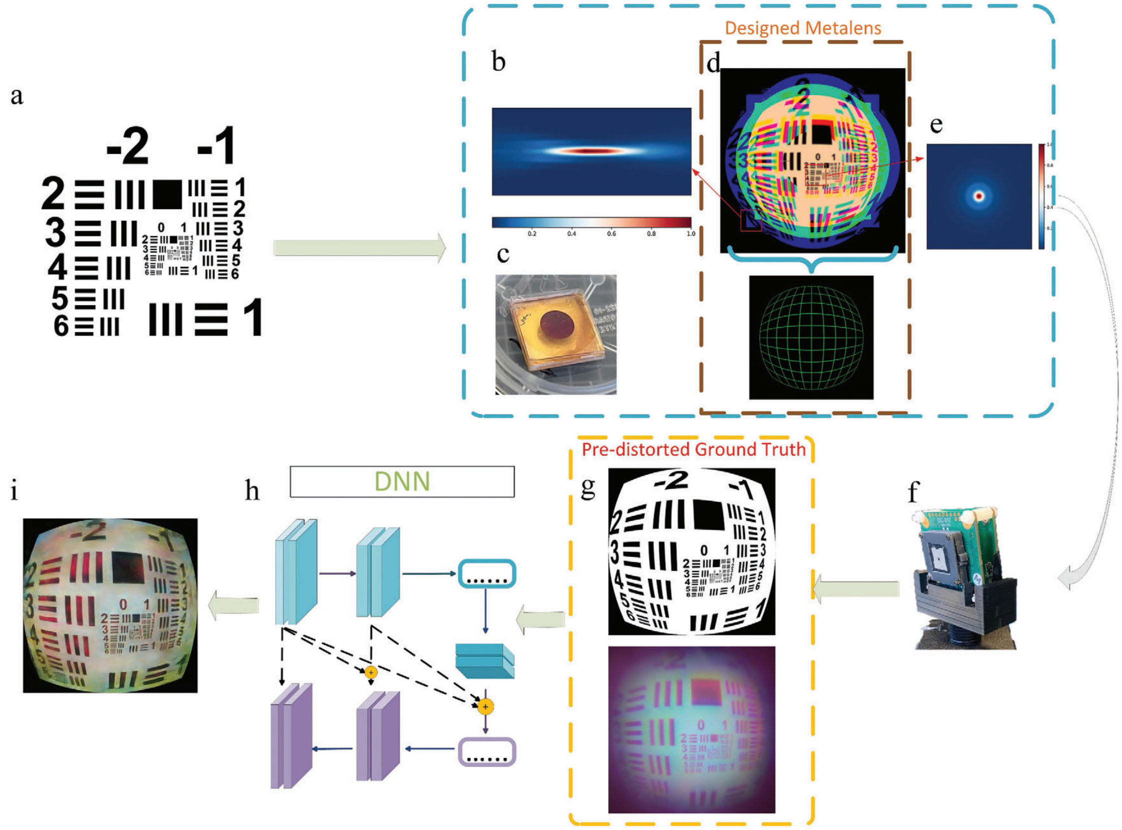

Figure 1. Overview of WFOV metalens imaging and restoration. a) USAF-1951 test chart used for simulation. b) 450–632 nm broadband PSF of metalens at 60-degree incidence. c) Photograph of fabricated metalens. d) Simulation result of USAF test chart and simulated distortion grid at 120-degree FOV. e) 450–632 nm broadband PSF of metalens at normal incidence. f) Photograph of the fabricated metalens mounted on a commercial camera. g) Raw image and pre-distorted ground truth image for DL model training. h) Proposed DL model diagram. i) Restoration results from the DL model.

applicability. Our results demonstrate that this model effectively corrects exposure, refocuses image and restores missing information on the raw captured images for Red-Green-Blue (RGB) channels. Additionally, MIRUnet can be extended to restore real-world images without extra training, further proving its ability to improve image fidelity across different metalens imaging platforms.

2. Results

2.1. Metalens Characteristics

In this study, the f/1.8 WFOV metalens (**Figure 1**c) was designed following the analytical approach we developed.[26] The metalens were fabricated on a Si-on-sapphire platform with a single-crystalline Si layer thickness of 600 nm. We note that even though single-crystalline Si is absorbing in the visible, the small metasurface thickness allows adequate light transmission throughout most of the visible band. The construction of the metalens includes an aperture stop (1.25 mm diameter) patterned on a separate glass substrate and subsequently bonded onto the metasurface substrate. The metalens is designed for monochromatic light at 525 nm, where it exhibits a near-180° FOV (more details can be found in Figure S1, Supporting Information). The polychromatic (for a flat band from 450 to 632 nm) PSFs at various incident angles are presented in Figure 1b,e. The lens exhibits limited aberration throughout the near-180° FOV at 525 nm. However, the severe impact of lateral chromatic aberration is evident from the elongated polychromatic PSFs at large incident angles (as shown in Figure 1b). The simulation results of the USAF chart in Figure 1d also demonstrate significant chromatic aberration in the visible spectrum. To address this issue, an image restoration algorithm was employed. A pre-distorted ground truth image (Figure 1g) and raw images from the metalens camera (Figure 1f) were used to train the Deep Neural Network (DNN) model (Figure 1h) to eliminate chromatic aberration (Figure 1i), which we shall discuss in detail in Section 2.2.

The optical performances of the WFOV metalens were experimentally assessed by imaging carefully selected test images displayed on an organic light-emitting diode (OLED) monitor. By capturing and analyzing photographs of these projected images, specific terms in the CITM can be quantified for the preprocessing techniques described in Section 2.2. **Figure 2**a illustrates the imaging resolution and chromatic aberrations from a USAF-1951 test chart with a white background taken by the WFOV metalens. In the center region (outlined in green), noticeable red and blue color fringes appear around the high-contrast edges of test patterns as a result of longitudinal chromatic aberration. It is apparent that the chromatic aberration increases toward the edges of the image with a clear offset between the

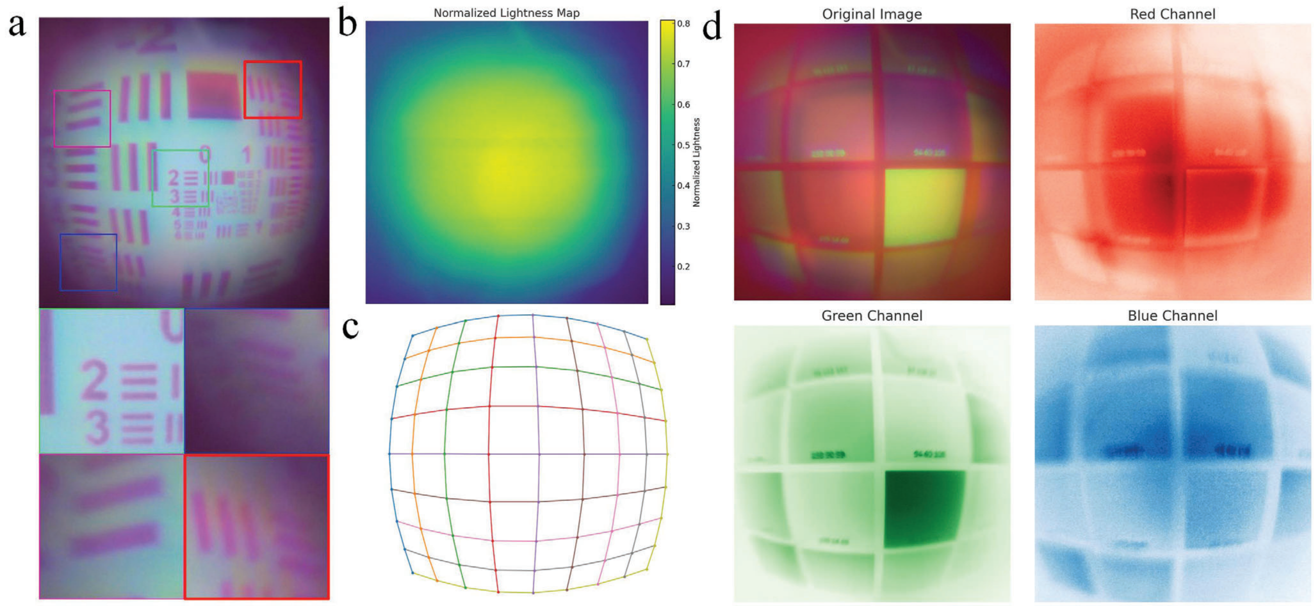

Figure 2. Optical performance of WFOV metalens. a) imaging resolution and chromatic aberration of a USAF-1951 test chart image. b) Lightness channel map of a HDR image. c) Distorted grid of the WFOV metalens image. d) Full-color test image from WFOV metalens and the corresponding red, green, and blue channels of test images.

color channels, indicative of the worsening lateral chromatic aberration which scales with the field angle. Moreover, the imbalance of intensity between the primary color channels due to the wavelength-dependent meta-atom transmission efficiency results in considerable color distortion.

Distortion is another optical aberration apparent in the WFOV images. Our theoretical analysis[26] has shown that a near-telecentric configuration is essential to minimizing other monochromatic aberrations. However, the image-space telecentricity leads to barrel distortion that monotonically increases toward the image edges at a rate much faster than non-telecentric optics (Figure 2c). This non-uniform distortion necessitates an extra pre-processing step for the image dataset used to train the DL model, where distortion matching that of the metalens is first applied to the undistorted ground truth images. This step is critical to ensuring that the DL model is trained on a coherent dataset to aptly counteract the distortion effect.

Another key factor that compromises the image quality is vignetting. Vignetting is an optical effect where the brightness and saturation in the periphery of an image are lower compared to the center. In our case, vignetting is primarily attributed to a cosine dependence of effective light-collecting aperture area of the metalens on the light incident angle. Vignetting can be quantified by analyzing the brightness distribution of the entire image. Figure 2b is the lightness channel map of a high dynamic range (HDR) synthesized image from a white test image on OLED monitor. The channel map demonstrates a strong vignetting effect, which leads to a brighter center region with more accurate color reproduction while the image periphery is darker and suffers from exacerbated color distortion. This imposes a serious challenge for color and resolution restoration, since loss functions of common deep learning-based algorithms rely on comparison between pixel values, and thus high-quality restoration is only viable with consistent feature extraction among whole images. Therefore, mitigating vignetting is also crucial for HDR imaging to restore uniform brightness and accurate color representation. A more elaborate comparison of HDR and Standard Dynamic Range (SDR) image brightness is given in Figures S2 and S3 (Supporting Information).

As a comprehensive example, Figure 2d shows a full-color test image captured by the metalens and the corresponding red, green, and blue channels. While the green channel is in reasonably sharp focus across the FOV, the red and blue channels are heavily aberrated. We also note that the object sizes projected on the image plane increase in the sequence of red-green-blue as a result of the lateral chromatic aberration.

2.2. CITM Induced Preprocessing, Raw Image Capturing Using HDR Synthesis

Achromatic WFOV metalens imaging systems impose serious difficulties associated with precise restoration of all color channels, complicated by off-axis monochromatic aberrations. For conventional metalens image restoration, some parameters like PSF are typically predefined to understand and correct image degradation. However, several problems arise with WFOV metalenses in such an approach. First of all, PSFs are dependent on the field angle and can be strongly affected by factors such as vignetting (Figure 2b) and distortion (Figure 2c). The vignette lowers the brightness of the image toward the image borders, while distortion changes the shape and accuracy of the image. All these effects change across the field of view, which makes PSF calibration a non-trivial task. Second, in WFOV metalenses systems, PSF rapidly changes at the edge of the image, making the measurements less accurate. This inaccuracy in the PSF model

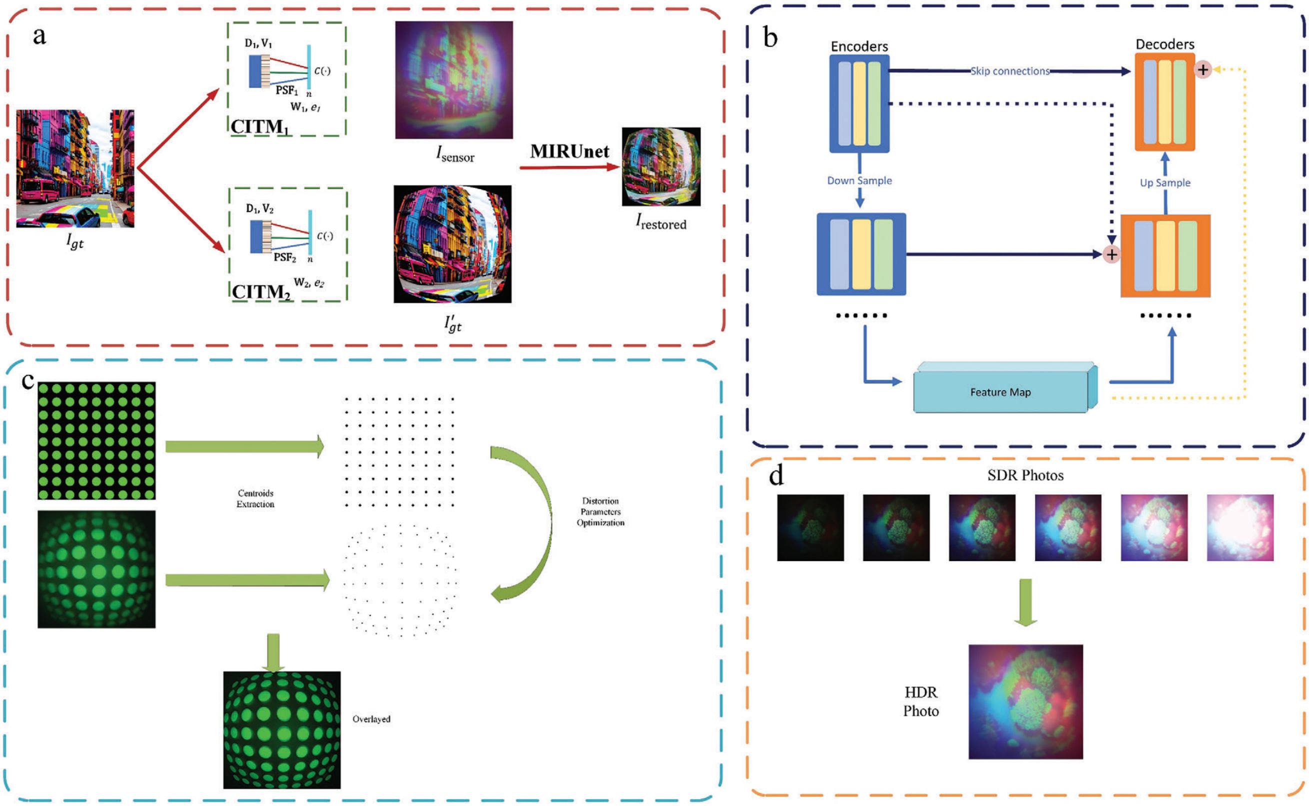

Figure 3. Processing process and DL model diagram. a) Distortion correction process using CITM. b) Diagram of MIRUnet. c) Pre-distortion process of ground truth image via centroids extraction. d) HDR synthesis process of raw image.

is going to result in errors in image reconstruction. A larger FOV means more regions of the image are under the influence of PSF variations, implying a more complex and finely tuned calibration process to take into account spatial variations across the whole FOV. Therefore, a large FOV metalens presents a exponentially more complicated case for image restoration.

For the purpose of relating captured images through metalens systems with their ground-truth counterpart, we have utilized the CITM[22] in our previous work. This relationship is important to design effective deep learning algorithms for image restoration. This is due to the inclusion of multiple common effects beside PSFs as expressed in Equation (1), where $I_{gt}\ (\lambda)$ is the wavelength component of the ground-truth image, T (·) is the transform function to form the ground-truth image to be captured, e.g., monitor projection. PSF(λ) is the effective PSF at each wavelength, *is for convolution, W is the color balance matrix, V is a factor to account for the vignetting effect, H is the homography matrix for perspective distortion, D (·) characterizes lens distortion, e is the exposure factor, n is the sensor noise, and C (·) sets the upper and lower bounds for signal clipping of the captured image.

$$I_{Sensor} = C\left(W \cdot V \cdot eD\left(T\left(I_{gt}(\lambda)\right)_{H^{-1}} * PSF(\lambda)\right) + n\right) \tag{1}$$

In this work, precautions were made before the data collection to simplify Equation (1). First, as described in our prior work, the monitor calibration process removes the term T(·) since the projected image is color-accurate with respect to the source files. Second, with due care in aligning the camera and the monitor, the homography matrix H can be discarded since the perspective errors that occur are negligible. In this respect, the simplified version of the CITM is given by:

$$I_{Sensor} = C\left(W \cdot V \cdot eDI_{gt}(\lambda)\ *PSF(\lambda) + n\right) \tag{2}$$

The remaining terms in Equation (2) are difficult to correct experimentally or numerically during data collection. Some terms, such as the distortion factor (D), color balance matrix (W), vignetting (V), and PSF(λ) are metalens' inherent properties, while the noise factor n is related to the camera sensor used. From the discussion in,[22] a robust DL model is capable of effectively restoring W, V, and PSF(λ) simultaneously, and the sensor noise n does not affect the restoration results. Thus, to apply MIRUNet to the WFOV metalens, the distortion factor D must be eliminated.

The distortion correction process can also be implemented using CITM. By setting up a virtual camera with a lens of given focal length, the ground truth image can be transformed to a new image with a certain amount of distortion without affecting other terms. The process can be explained using CITM by incorporating a distortion term directly to $I_{gt}\ (\lambda)$, as in Equation (3). Multiple iterations using Equation (3) are needed to approach the distortion profile. As shown in **Figure** 3c, an iterative optimization process is used to better fit the true distortion profile, and the details can be found in S2.1 (Supporting Information).

$$I'_{gt} = D' \cdot I_{gt} \tag{3}$$

After the optimization, since the $D' \approx D_{gt}$, (Equations (2) and (3) can be combined, and the captured image through WFOV metalens system is then represented as:

$$I_{sensor} = C\left(W \cdot V \cdot eDI'_{gt} * PSF(\lambda) + n\right) \tag{4}$$

where I'_{gt} is the pre-distorted ground truth image as shown in Figure 3a. Equation (4) reveals that the image restoration problem for WFOV metalenses is equivalent to that of a standard metalens with a smaller FOV. Consequently, MIRUNet should be capable of handling the inverse of the remaining terms in the equation to enhance the overall image quality.

Besides the distortion correction, the lack of brightness in the periphery of the captured image requires further processing to make both the brightest and darkest areas within the FOV free from the clipping function C (·). Figure 3d shows several photographs taken with different exposure time. Clearly, no SDR photo provides a good result in terms of balanced brightness in the entire frame of the image. Therefore, the need for HDR synthesis becomes a necessity for achieving the required FOV. Figure 3d also demonstrates the output of the HDR synthesis process. While an SDR photo can reveal details around the edges of a single photo when increasing the exposure duration, the center region of the photo is overexposed. For an HDR photo, the exposure levels of the image are already set and details across the whole frame remains visible. Therefore, the exposure factor e becomes constant, and the clipping function C (·) can be ignored, Equation (4) becomes:

$$I_{sensor} = W \cdot V \cdot DI'_{gt} * PSF(\lambda) + n \tag{5}$$

With fewer factors to correct in Equation (5) compared to the original CITM (Equation 1), the deep learning network can achieve better performance and generalization by mostly focusing on the metalens-related terms. This ensures the network's proper function outside of the training setup. The detailed description of HDR synthesis can be found in S2.2 (Supporting Information).

2.3. DNN Formation and Image Reconstruction Results

The preceding discussions point out that a sophisticated image reconstruction algorithm is necessary to reverse the CITM forward transformation to recover captured images through metalens systems. By inverting the effects of ill-informed PSFs, vignetting, and color balance, high-resolution full-color images can be expected, however such process is usually impossible using standard analytical methods. To handle this issue, we have developed an advanced approach[22] —namely, the "MIRUnet", which effectively compensates for most of the nonlinear factors in the CITM through a single process step without measuring and quantifying explicit parameters.

MIRUnet is a U-Net[27,28] architecture variant with multiscale capabilities, specially designed for the metalens image enhancement and reconstruction tasks. In the main structure of the MIRUnet model, with an encoder and a decoder, additional skip connections and an attention module are attached to refine feature selection and increase learning efficiency. Especially, WFOV metalenses suffer more from the negative effects of inconsistent PSFs across the image plane, vignetting, and dispersion-related transmission efficiency compared with metalenses with a smaller FOV. Also, the distorted ground truth images left black borders which could affect the calculation of peak signal-to-noise ratio (PSNR) as shown in **Figure 4**. Therefore, we optimized the original MIRUnet model to achieve better feature extraction to handle this challenging problem. First, weightage to PSNR loss was increased, with the capability of recovering image quality at the basic feature level.[29] Second, we replaced the Leaky Rectified Linear Unit (ReLU) activation function with Gaussian Error Linear Unit (GELU) to allow smooth and adaptive nonlinear transformation that can deal with the changes in the image features more accurately.[30] We also combined the output of the last decoder with the original input image via element-wise addition, for the network to directly incorporate the learned modifications with the original input, highlighting both the learned features and the original image details.[31] The modified MIRUnet network diagram can be seen in Figure 3b. With a fully trained network, we could restore an image in less than 0.15 s. It is obvious that there is room to improve our model to achieve real-time image processing at 30- frames per second (fps). We believe that future optimizations, such as utilizing specialized hardware or applying network pruning techniques, could effectively address these limitations and give MIRUnet real-time processing capabilities.

Training of the modified MIRUnet model was conducted on a diversified dataset created by generative deep learning methods such as Stable Diffusion.[32] This synthetic dataset was created to ensure a sufficient variation in colors, contrast, brightness, and objects that would represent the possible conditions for the imaging of the metalens in a real environment.[33] Barrel distortion, commonly encountered in WFOV imaging, was added to the images in the dataset through an optimized distortion term derived from the method shown in Figure 3c. The model was forced to become highly adaptable to various optical conditions and predict the random nature of real-world optical imperfections with high accuracy and reliability. The model was trained over 100 epochs using the Adam optimizer,[34] utilizing the computational power of two Nvidia Geforce RTX 4090 graphics processing units (GPUs). The size of input/output images is set to 512 × 512 pixels to guarantee efficient handling and processing of the data for optimal learning and better image restoration. After the configuration of the experimental setup, 200 synthetic images were used for evaluation.

Comparative analysis of different image restoration models proves that MIRUNet considerably outperforms other models in restored image quality. In Figure 4, the performance of our model is contrasted with two other commonly used methods: Multi Scale Retinex with Color Restoration (MSRCR) [35] which is traditionally used for restoring resolution and color, and Revitalizing Real Image Dehazing via High-Quality Codebook Priors (RIDCP),[36] the state-of-the-art algorithm for dehazing. Four synthetic images are selected in Figure 4 to check the

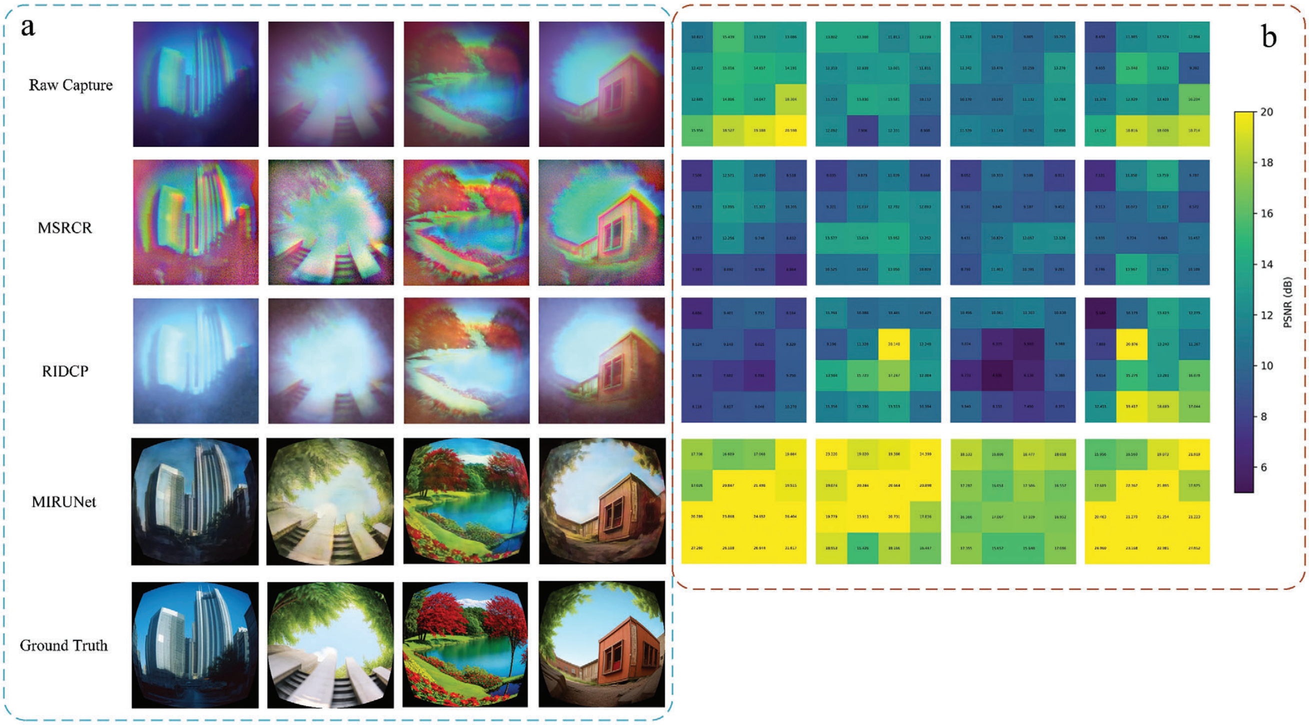

Figure 4. Image recovery results of MIRUnet compared with results from state-of-the-art methods. a) Comparison between the MIRUnet output and other methods was conducted on four types of images: one under-exposed, one over-exposed, one with an RGB color scheme, and one with a CMY color scheme. b) Pixel maps of corresponding recovery outputs.

performance of restoration algorithms under different scenarios and different lighting conditions and color palettes. These images include one under-exposed, one over-exposed, one with an RGB color scheme, and one with a CMY color scheme, respectively. The ground truth images in Figure 4 are pre-distorted computer generate images. After pre-distortion, the synthetic images were displayed on an OLED screen then captured by the WFOV metalens camera as raw images, which were used as inputs for the MIRUNet and other state-of-the-art models.

Figure 4a shows that visually, the raw capture images have chromatic aberrations and blurring. The MSRCR method partially enhances the clarity by boosting the brightness of dark regions and increases the contrast across the image but does so with noticeable color artifacts and added noise. The RIDCP model improves clarity overall but maintains high level of aberrations and blurring, along with inaccurate color restoration with an overexposed appearance. In comparison, MIRUNet greatly enhances the quality of the images, replicating the ground truth images with minimal distortions, evenly distributed brightness, and accurate color reproduction. It is clear that MIRUNet mitigates chromatic aberrations and provides improved focus uniformity across the full range of FOV.

Numerically, Figure 4b shows the PSNRs of each restored image with respect to its ground truth. MSRCR and RIDCP did not produce an observable PSNR increase, and in most regions, the PSNR even deteriorated due to less accurate color representation compared to raw captures. In stark contrast, MIRUNet demonstrates significant and uniform PSNR increases across the entire image, including corners where vignetting, distortion and chromatic aberration are most severe, an outcome that neither of the other methods can produce. Besides, MIRUNet also keeps a high PSNR in regions of over- or under-exposure, which indicates its excellent capability of exposure correction and dynamic range extension. We also note that raw capture images show higher PSNR values in the corners than in the center because of the combination of underexposure conditions with the black borders of ground truth images. On the other hand, for the restored images using RIDCP, the second column of Figure 4a, some peak PSNRs are introduced into the center because pixels are saturated in these bright areas, and the ground truth image also shares the near maximum pixel values.

3. Discussion

To further evaluate the efficiency of MIRUNet and CITM-based preprocessing, an experiment with a flexible OLED display showing real-world scenarios was conducted. The experimental environment is set up such that the flexible display is encapsulated between two 3D-printed holders whose bending radius is equal to half the length of the longer edge of the OLED screen. The two holders and the OLED screen can be flipped. As an effective augmentation method, image flipping is widely used to extend the variability of deep learning training set.[37] Similarly, the proper restoration of flipped image can also be used as a measure to ensure MIRUNet is not over-trained. The OLED screen is bright enough to be reflected by the shining plastic holders, making the reflection observable by the camera when increasing the exposure duration, which can also be observed in the MIRUNet

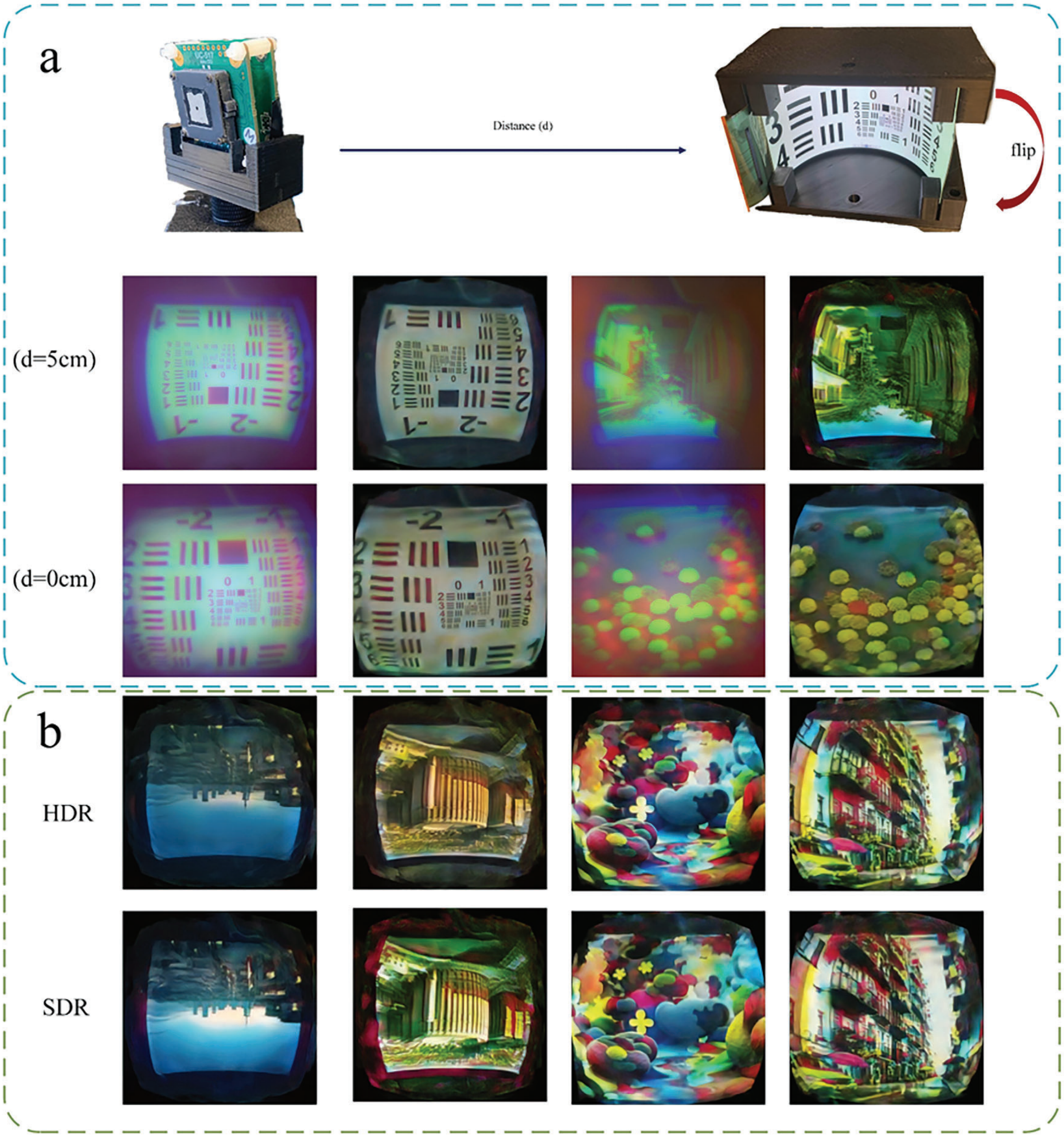

Figure 5. Image recovery results of MIRUnet in a real-world scenario. a) Recovery of images via pre-trained MIRUnet captured at different distances. b) Recovery of HDR and SDR images via pre-trained MIRUnet.

restored images. The high brightness of the OLED screen and reflections from the 3D-printed holder surpassed the network's restoration capabilities, causing unwanted artifacts around the image in **Figure 5**.

To explore and expand the utility of the WFOV metalens imaging system under various conditions, this experiment was conducted to test the system at shorter focusing distances. Although the lens is designed for infinity focus, the following results (as shown in Figure 5) aim to demonstrate its potential flexibility for applications beyond its intended design, such as in close-up imaging scenarios. Figure 5a shows the restoration performance of MIRUNet on images captured at distances of d = 5 cm and d = 0 cm (with the metalens situated at the center of the circular holder). Since these distances are not optimal for the lens's operation, the images captured exhibit different color fringes and distortions, presenting additional challenges for the restoration process.

Notice that, at a distance of 5 cm, the raw images suffer from noticeable chromatic aberrations and blurring. On the other hand, after restoration with MIRUNet, the chromatic aberration is largely mitigated, and the clarity is restored closer to that of the original test patterns. The region outside of the OLED screen remains dark after restoration. At a distance of 0 cm, compared with the previous case, the distortion across the image has changed dramatically, which also leads to different cases of color fringes and even focusing. Regardless of the complex image projection scheme, MIRUNet amends these degradations and produces images with higher fidelity, better sharpness, and color accuracy. It is worth mentioning that even at different distances than those in the training set, MIRUNet processes the raw captures well, demonstrating its robustness and adaptability. Thus, MIRUNet is a versatile method that can be applied to different object distances, ensuring that it performs consistently in different situations.

Another comparison is given in Figure 5b between HDR and SDR images, which shows how these formats affect performance in the real world. On the HDR image, the brightness is distributed more widely in the range, which captures the restorations clearer and much more detailed. The restoration is clearest in the top row of Figure 5b. In contrast, MIRUNet was not specifically trained on SDR images; it still makes consistent restoration of SDR images although with increased noise levels and reduced detail, especially in regions with large variations of brightness. Some regions with severe under- or over-exposure appear slightly color shifted, but the details as well as geometry of the restored images remain consistent. Nevertheless, the higher noise in SDR restorations highlights the importance of HDR in real-world applications, which concerns image quality over varying lighting conditions. This makes it important to use CITM analysis on the imaging systems to recognize the key factors that affect the performance of such systems.

4. Conclusion

In this study, we present a novel approach to achieve WFOV achromatic metalens imaging through the application of deep learning techniques. By integrating an WFOV metalens system, CITM, and the MIRUnet DL model, the research effectively addresses major metalens imaging performance degradation issues such as chromatic aberration, vignetting, and distortion. The experimental results demonstrate that the proposed method significantly enhances image quality with accurate color reproduction, uniform brightness, and high resolution across a large FOV. The findings underscore the potential of combining metalens technology with deep learning to overcome inherent optical limitations, paving the way for practical applications in advanced imaging systems.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

This work was partially funded by the DARPA ENVision program (Grant # HR00112290005).

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

deep learning, Imaging, metalens, meta-optics, neural network, wide field-of-view

Received: August 16, 2024
Revised: December 2, 2024
Published online: December 25, 2024

[1] W. T. Chen, A. Y. Zhu, J. Sisler, Z. Bharwani, F. Capasso, *Nat. Commun.* **2019**, *10*, 355.
[2] M. Khorasaninejad, Z. Shi, A. Y. Zhu, W. T. Chen, V. Sanjeev, A. Zaidi, F. Capasso, *Nano Lett.* **2017**, *17*, 1819.
[3] W. T. Chen, A. Y. Zhu, F. Capasso, *Nat. Rev. Mater.* **2020**, *5*, 604.
[4] S. S. Bukhari, J. (Yiannis) Vardaxoglou, W. Whittow, *Appl. Sci.* **2019**, *9*, 2727.
[5] M. Y. Shalaginov, S. An, F. Yang, P. Su, D. Lyzwa, A. M. Agarwal, H. Zhang, J. Hu, T. Gu, *Nano Lett.* **2020**, *20*, 7429.
[6] F. Yang, M. Y. Shalaginov, H.-I. Lin, S. An, A. Agarwal, H. Zhang, C. Rivero-Baleine, T. Gu, J. Hu, *Adv. Photon.* **2023**, *5*, 033001.
[7] F. Yang, T. Gu, J. Hu, *J. Opt. Soc. Am. B* **2023**, *40*, 2211.
[8] A. Martins, J. Li, B.-H. V. Borges, T. F. Krauss, E. R. Martins, *Nanophotonics* **2022**, *11*, 1187.
[9] B. Groever, W. T. Chen, F. Capasso, *Nano Lett.* **2017**, *17*, 4902.
[10] X. Luo, F. Zhang, M. Pu, Y. Guo, X. Li, X. Ma, *Nanophotonics* **2022**, *11*, 0583.
[11] A. Arbabi, E. Arbabi, S. M. Kamali, Y. Horie, S. Han, A. Faraon, *Nat. Commun.* **2016**, *7*, 13682.
[12] H. Liang, A. Martins, B.-H. V. Borges, J. Zhou, E. R. Martins, J. Li, T. F. Krauss, *Optica* **2019**, *6*, 1461.
[13] S. Shrestha, A. C. Overvig, M. Lu, A. Stein, N. Yu, *Light: Sci. Appl.* **2018**, *7*, 85.
[14] S. Wang, P. Wu, V.-C. Su, Y.-C. Lai, M.-K. Chen, H. Kuo, B. H. Chen, Y. Chen, T.-T. Huang, J.-H. Wang, R.-M. Lin, C.-H. Kuan, T. Li, Z. Wang, S. Zhu, D. P. Tsai, *Nat. Nanotechnol.* **2018**, *13*,.
[15] W. T. Chen, A. Y. Zhu, V. Sanjeev, M. Khorasaninejad, Z. Shi, E. Lee, F. Capasso, *Nat. Nanotechnol.* **2018**, *13*, 220.
[16] F. Presutti, F. Monticone, *Optica* **2020**, *7*, 624.
[17] K. Shastri, F. Monticone, *EPJ Appl. Metamat.* **2022**, *9*, 16.
[18] F. Yang, S. An, M. Y. Shalaginov, H. Zhang, C. Rivero-Baleine, J. Hu, T. Gu, *Opt. Lett.* **2021**, *46*, 5735.
[19] J. Engelberg, U. Levy, *Opt. Express* **2017**, *25*, 21637.
[20] Y. Liu, J. Zhang, X. L. Roux, E. Cassan, D. Marris-Morini, L. Vivien, C. Alonso-Ramos, D. Melati, *Opt. Express* **2022**, *30*, 39860.
[21] A. I. Kuznetsov, M. L. Brongersma, J. Yao, M. K. Chen, U. Levy, D. P. Tsai, N. I. Zheludev, A. Faraon, A. Arbabi, N. Yu, D. Chanda, K. B. Crozier, A. V. Kildishev, H. Wang, J. K. W. Yang, J. G. Valentine, P. Genevet, J. A. Fan, O. D. Miller, A. Majumdar, J. E. Fröch, D. Brady, F. Heide, A. Veeraraghavan, N. Engheta, A. Alù, A. Polman, H. A. Atwater, P. Thureja, R. Paniagua-Dominguez, et al., *ACS Photonics* **2024**, *11*, 816.
[22] Y. Dong, B. Zheng, H. Li, H. Tang, H. Zhao, Y. Huang, S. An, H. Zhang, *ACS Photonics* **2024**, *11*, 1645.
[23] R. Maman, E. Mualem, N. Mazurski, J. Engelberg, U. Levy, *ACS Photonics* **2023**, *10*, 4494.
[24] J. Seo, J. Jo, J. Kim, J. Kang, C. Kang, S. Moon, E. Lee, J. Hong, J. Rho, H. Chung, *Adv. Photon.* **2024**, *6*, 066002.
[25] Y. Liu, W.-D. Li, K.-Y. Xin, Z.-M. Chen, Z.-Y. Chen, R. Chen, X.-D. Chen, F.-L. Zhao, W.-S. Zheng, J.-W. Dong, *Adv. Photon.* **2024**, *6*, 056001.
[26] F. Yang, S. An, M. Y. Shalaginov, H. Zhang, J. Hu, T. Gu, *Chin. Opt. Lett.* **2023**, *21*, 023601.
[27] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, J. Liang, in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical*

Decision Support (Eds.: D. Stoyanov, Z. Taylor, G. Carneiro, T. Syeda-Mahmood, A. Martel, L. Maier-Hein, J. M. R. S. Tavares, A. Bradley, J. P. Papa, V. Belagiannis, J. C. Nascimento, Z. Lu, S. Conjeti, M. Moradi, H. Greenspan, A. Madabhushi), Springer International Publishing, Berlin, Germany, **2018**, pp. 3–11.

[28] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y. Chen, J. Wu, ICASSP 2020 –2020 IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, Piscataway, NJ **2020**, 1055.

[29] H. Zhao, O. Gallo, I. Frosio, J. Kautz, *IEEE Trans. Computational. Imaging*. **2017**, *3*, 47.

[30] D. Hendrycks, K. Gimpel, arXiv **2016**, arxiv.org/abs/1606.08415.

[31] L. Chen, X. Chu, X. Zhang, J. Sun, in *Computer Vision – ECCV 2022*, (Eds: S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, T. Hassner), Springer Nature, Switzerland **2022**, pp. 17–33.

[32] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, B. Ommer, in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, New Orleans, LA, USA **2022**, pp. 10674–10685.

[33] X. Wang, L. Xie, C. Dong, Y. Shan, in *2021 IEEE/CVF Int. Conf. on Computer Vision Workshops (ICCVW)*, IEEE, Montreal, BC, Canada **2021**, pp. 1905–1914.

[34] D. Kingma, J. Ba, arXiv **2014**, arxiv.org/abs/1412.6980.

[35] A. B. Petro, C. Sbert, J.-M. Morel, *Image Processing On Line* **2014**, *4*, 71.

[36] R.-Q. Wu, Z.-P. Duan, C.-L. Guo, Z. Chai, C.-Y. Li, in 2023 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), IEEE, Vancouver, BC, Canada **2023**, pp. 22282–22291.

[37] K. Simonyan, A. Zisserman, in 3rd Int. Conf. on Learning Representations (ICLR 2015) , *Computational And Biological Learning Society* **2015**, pp. 1–14.