Next-Generation Metalens Vision System: Powered by AI and Applied to AI

Fen Fang[1], **Muli Yang**[1*], **Henan Wang**[2], **Xinan Liang**[3,4], **Tobias Mass**[5], **Xuewu Xu**[3,4], **Xulei Yang**[1], **Zhengguo Li**[1]

[1]Institute for Infocomm Research (I[2]R), A*STAR, Singapore
[2]Independent Researcher
[3]Institute of Materials Research and Engineering (IMRE), A*STAR, Singapore
[4]National Semiconductor Translation and Innovation Center (NSTIC), Singapore
[5]MetaOptics Technologies, Singapore
{fangf, yangml, liangx, yangx, xux1, ezgli}@a-star.edu.sg, nnhhwang@gmail.com, tobias@metaoptics.sg

Abstract

Metalenses have been widely recognized as a key building block of next-generation optical systems, offering unprecedented advantages in compactness, lightweight design, and scalable manufacturing compared to traditional refractive optics. Despite this promise, practical use is limited by optical aberrations, blur, and illumination sensitivity, which degrade both visual quality and machine perception. In this demonstration, we present an *end-to-end metalens vision system*—from hardware sensing with a custom-built RGB metalens camera, to physics-informed imaging and real-time restoration, and finally to downstream vision applications such as object detection and depth estimation. By integrating spatially-aware attention enhancement and reinforcement learning-based illumination control into a real-time system, our solution transforms degraded raw captures into high-fidelity images that are both visually interpretable and functionally reliable for machine vision. This AI-powered pipeline highlights metalenses as a cornerstone for next-generation imaging, where advances in optics and machine intelligence jointly drive the future of visual perception.

Website — https://metalens-star.github.io

Introduction

Conventional optical systems, though mature, remain bulky, heavy, and costly to manufacture, limiting their deployment in compact platforms such as mobile devices, wearables, and biomedical systems. Metalenses offer an attractive alternative with their ultra-thin design, multifunctionality, and compatibility with electronic integration (Khorasaninejad and Capasso 2017), and are increasingly recognized as a key enabler for next-generation imaging and sensing.

However, metalens imaging still suffers from optical aberrations (Zheng et al. 2022), chromatic distortions (Smith et al. 2020), and peripheral blur (Lin et al. 2021), which severely degrade both perceptual quality and machine vision reliability. Hardware remedies (e.g., multi-layer metasurfaces) reduce these distortions but incur high fabrication costs (Huang et al. 2019), while generic software-based restoration struggles with the highly spatially varying, illumination-sensitive degradations unique to metalenses (Johnson et al. 2021).

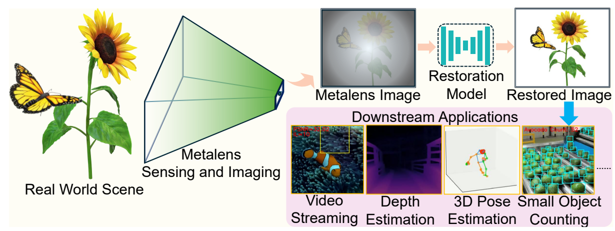

Figure 1: Our proposed metalens vision system with four end-to-end stages: *sensing*, *imaging*, *restoration*, and various *downstream applications*. [scriptsize: Real World Scene]

As illustrated in Fig. 1, we address this gap with a *real-time, AI-driven pipeline* tailored for metalens imaging. A raw metalens image is restored into a high-quality representation, benefiting both human perception and downstream application. Our system integrates physics-informed priors with deep learning, leveraging spatially-aware attention to recover peripheral details and a reinforcement learning–based module to adapt robustly to unseen lighting conditions. By integrating hardware sensing, metalens imaging, algorithmic restoration, and downstream application into an end-to-end pipeline, our system shows how AI-powered imaging can fully realize the potential of metalenses.

System Overview

We present an *end-to-end AI-powered metalens imaging system solution* that spans the entire pipeline from raw sensing to intelligent applications:

- **Metalens Sensing and Imaging.** Raw RGB images are acquired directly from a custom-built metalens camera prototype with a wide field of view (FoV) of 70 (*vs.* prevalent FoV of 30 on the market). The system uses three single-layer metalenses fabricated on the same substrate, each designed for red, green, and blue wavelengths. These captures are aligned and fused to form

*Corresponding author.

Figure 2: User interface of our demo. In this figure, we illustrate the *small object counting* task using *raw*, *ground truth*, and *restored metalens videos*, respectively.

a color image, faithfully preserves real-world degradations, including aberrations, chromatic distortions, peripheral blur, and illumination sensitivity, which provide authentic inputs that cannot be replicated by simulations.

- **Real-Time Image Restoration.** At the core of our system is a hybrid restoration engine that combines traditional techniques—namely geometric alignment (Bao, Yang, and Jin 2014), intensity normalization, chromatic aberration correction (Sun et al. 2017), and deblurring (Nah, Kim, and Lee 2017)—with our proposed deep learning-based approach, detailed in the next section. Together, these components form a unified real-time pipeline, capable of processing 1 MP frames at an average speed of 46 ms per frame on a single GPU.
- **Downstream Vision Applications.** – The restored outputs are directly consumed by multiple AI tasks, including high-fidelity video streaming (Seo et al. 2024; Chu, Chen, and Yu 2022), depth estimation (Chen et al. 2025), 3D pose estimation (Jiang et al. 2023), and small object counting (Jocher, Chaurasia, and Qiu 2023). Our user interface is shown in Fig 2. During the live demo session, participants will observe side-by-side comparisons of raw and restored streams under dynamic lighting, and experience measurable gains in task-level accuracy (e.g., higher detection mAP and more stable depth maps). This setup highlights both the visual appeal and the functional value of our proposed system.

The key characteristics of our proposed system are:

- **Speed.** Processes 1 MP images in 46 ms on a single RTX 3090 GPU.
- **Adaptability.** Supports different metalens prototypes and Point Spread Function (PSF) models without the necessity of retraining.
- **Scalability.** Modular design enables seamless integration into existing imaging workflows.
- **Versatility.** Enhances both human visual experience and AI pipelines, bridging optics and perception.

From Demo to Product. While demonstrated as a research prototype, this pipeline points toward a future where hardware metalenses, imaging systems, restoration algorithms, and computational platforms are *co-designed and fused into a single product*. Such a system can function as a complete AI-empowered imaging device, or be modularized into software solutions for integration into broader vision pipelines. In both cases, it highlights the potential of metalens-based systems to become a cornerstone of next-generation optical imaging technologies.

Methodology

Our system's real-time restoration framework converts raw metalens captures into high-quality images for human and machine use, showcasing two key innovations.

Spatial-Aware Attention Scheduler. To effectively restore wide-field-of-view metalens images, we design a spatial-aware attention scheduler that allocates computational resources according to regional degradation severity. Peripheral regions, which typically suffer from stronger blur and aberrations (Tseng et al. 2021), receive higher attention weights, while central areas are processed with lighter computation. This adaptive scheduling mechanism not only improves restoration fidelity in challenging regions but also reduces redundant computation, enabling real-time performance without sacrificing quality.

Retraining-Free Illumination Adapter. Metalens imaging is highly sensitive to illumination changes, which can severely degrade restoration quality if left unaddressed. We propose a reinforcement learning (RL) (Sutton and Barto 2018)–based illumination adapter that operates at inference time to dynamically adjust the input illumination profile. Rather than retraining the restoration model under each lighting condition, a lightweight RL agent learns to modify illumination on-the-fly, guided by a reward signal derived from restoration quality. This retraining-free strategy ensures stable, reliable, and generalizable performance across diverse lighting scenarios, without adding any significant computational overhead.

Together, these two innovations form the algorithmic core of our system: the attention scheduler enhances spatial adaptivity and efficiency, while the RL-based adapter ensures illumination robustness. By unifying these components, our pipeline delivers high-quality, real-time restoration that seamlessly integrates into downstream vision tasks.

Conclusion

In this demo, we showcase a complete end-to-end metalens vision system that seamlessly integrates hardware innovation with intelligent computation. Our pipeline combines a custom-built RGB metalens camera with physics-informed imaging, real-time restoration, and reinforcement learning–based illumination adaptation, enabling the transformation of degraded raw captures into high-fidelity images suitable for both human viewing and machine vision tasks. Collectively, these results highlight metalenses—with their unparalleled advantages in compactness, lightweight design, and scalable manufacturing—as a pivotal technology for future imaging, where advances in optics and AI converge to transform the way we perceive the world.

Acknowledgements

This work was supported by the Agency for Science, Technology and Research (A*STAR) under its GAP project (Grant No. I24D1AG062) and its MTC Programmatic Funds (Grant No. M23L7b0021).

References

Bao, L.; Yang, Q.; and Jin, H. 2014. Geometric Image Alignment by Learning. In *IEEE Conference on Computer Vision and Pattern Recognition*.

Chen, S.; Guo, H.; Zhu, S.; Zhang, F.; Huang, Z.; Feng, J.; and Kang, B. 2025. Video Depth Anything: Consistent Depth Estimation for Super-Long Videos. *arXiv preprint arXiv:2501.12375*. Accepted at CVPR 2025 Highlight.

Chu, X.; Chen, L.; and Yu, W. 2022. NAFSSR: Stereo Image Super-Resolution Using NAFNet. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 1239–1248.

Huang, Y.; et al. 2019. Multi-layer metasurface design for aberration correction. In *Optica*.

Jiang, T.; Lu, P.; Zhang, L.; Ma, N.; Han, R.; Lyu, C.; Li, Y.; and Chen, K. 2023. RTMPose: Real-Time Multi-Person Pose Estimation based on MMPose. arXiv:2303.07399.

Jocher, G.; Chaurasia, A.; and Qiu, J. 2023. Ultralytics YOLOv8. https://github.com/ultralytics/ultralytics. Accessed: 2025-09-12.

Johnson, R.; et al. 2021. Generic software restoration methods for computational imaging. In *ICCV*.

Khorasaninejad, M.; and Capasso, F. 2017. Metalenses: Versatile multifunctional photonic components. *Science*, 358(6367): eaam8100.

Lin, J.; et al. 2021. Characterizing aberrations in metalens imaging. In *CVPR Workshops*.

Nah, S.; Kim, T. H.; and Lee, K. M. 2017. Deep Multi-scale Convolutional Neural Network for Dynamic Scene Deblurring. In *IEEE Conference on Computer Vision and Pattern Recognition*.

Seo, J.; Jo, J.; Kim, J.; Kang, J.; Kang, C.; Moon, S.; Lee, E.; Hong, J.; Rho, J.; and Chung, H. 2024. Deep-learning-driven End-to-End Metalens Imaging. *Advanced Photonics*, 6(6): 066002.

Smith, A.; et al. 2020. Chromatic distortions in metasurface lenses. *Optics Express*, 28: 12345–12358.

Sun, T.; et al. 2017. Chromatic Aberration Correction via Spectral Domain Alignment. In *Proceedings of the IEEE International Conference on Computer Vision*.

Sutton, R. S.; and Barto, A. G. 2018. *Reinforcement Learning: An Introduction*. MIT Press, 2nd edition.

Tseng, E.; Colburn, S.; Whitehead, J.; Huang, L.; Baek, S.-H.; Majumdar, A.; and Heide, F. 2021. Neural nano-optics for high-quality thin lens imaging. *Nature Communications*, 12: 6493.

Zheng, G.; et al. 2022. Deep learning-based computational imaging through metalenses. *Nature Communications*, 13: 4567.