

Fovea Stacking: Imaging with Dynamic Localized Aberration Correction

SHI MAO, King Abdullah University of Science and Technology (KAUST), Saudi Arabia

YOGESHWAR NATH MISHRA*, King Abdullah University of Science and Technology (KAUST), Saudi Arabia

WOLFGANG HEIDRICH, King Abdullah University of Science and Technology (KAUST), Saudi Arabia

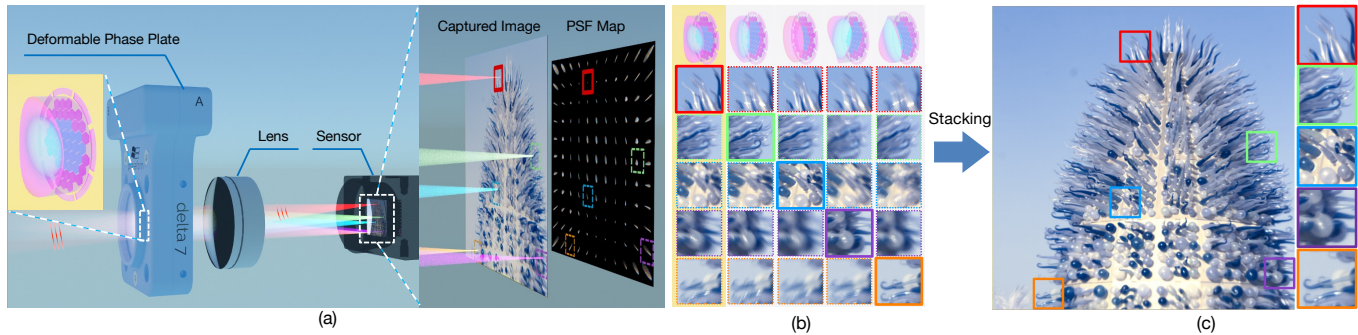


Fig. 1. Fovea Stacking is a novel computational camera paradigm aimed at simplifying (and ultimately miniaturizing) camera optics. (a) The optical system consists of an achromatic doublet lens, which is highly aberrated, and a refractive deformable phase plate (DPP), which can be used to correct the aberrations for a *local* region of interest (the fovea). While this optical system cannot simultaneously correct all aberrations across the entire image, the DPP is capable of dynamically moving the fovea *anywhere* in the image. For example, the DPP deformation in (a) corrects aberrations along the red beam for the corresponding red box fovea on the sensor. (b) shows magnified crops of foveated images captured with different DPP deformations. The left column shows the image from (a), the other columns have the fovea moved to the other respective boxes. (c) By stacking up the sharpest regions from different foveated images (i.e. a fovea stack), an un-aberrated all-in-focus image can be recovered. Other applications such as object tracking and depth-dependent focusing are discussed in the text.

The desire for cameras with smaller form factors has recently led to a push for exploring computational imaging systems with reduced optical complexity such as a smaller number of lens elements. Unfortunately such simplified optical systems usually suffer from severe aberrations, especially in off-axis regions, which can be difficult to correct purely in software.

In this paper we introduce Fovea Stacking, a new type of imaging system that utilizes an emerging dynamic optical component called the deformable phase plate (DPP) for localized aberration correction anywhere on the image sensor. By optimizing DPP deformations through a differentiable optical model, off-axis aberrations are corrected locally, producing a foveated image with enhanced sharpness at the fixation point - analogous to the eye's fovea. Stacking multiple such foveated images, each with a different fixation point, yields a composite image free from aberrations. To efficiently cover the entire field of view, we propose joint optimization of DPP deformations under imaging budget constraints. Due to the DPP device's non-linear behavior, we introduce a neural network-based control model for improved agreement between simulation and hardware performance.

We further demonstrated that for extended depth-of-field imaging, Fovea Stacking outperforms traditional focus stacking in image quality. By integrating object detection or eye-tracking, the system can dynamically adjust the lens to track the object of interest-enabling real-time foveated video

suitable for downstream applications such as surveillance or foveated virtual reality displays.

CCS Concepts: • **Computing methodologies** → **Computational photography**.

Additional Key Words and Phrases: Computational Imaging, Aberration Correction, Differentiable Optics, Image Fusion

ACM Reference Format:

Shi Mao, Yogeshwar Nath Mishra, and Wolfgang Heidrich. 2025. Fovea Stacking: Imaging with Dynamic Localized Aberration Correction. *ACM Trans. Graph.* 1, 1 (September 2025), 13 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

1 INTRODUCTION

To achieve high image quality across the full field of view (FoV), conventional imaging systems use complex lens systems that are carefully designed to correct for optical aberrations both on- and off-axis. Great strides have been taken to miniaturize such designs to achieve impressive imaging results in small form factors, for example in mobile devices. Unfortunately further miniaturization is approaching physical limits with existing design philosophies and optical components such as refractive lenses.

New opportunities arise from the emergence of new types of dynamically tunable optical components that facilitate completely new design paradigms. The most well-known type of such elements are probably liquid tunable lenses [Liu et al. 2023]. Initially bulky and expensive, this technology has recently become small and affordable enough to be used in mobile phones [Blain 2021]. A more recent type of dynamically tunable optical component is the deformable phase plate (DPP) [Banerjee et al. 2018; Rajaeipour et al. 2021].

*Now at Indian Institute of Technology Jodhpur

Authors' addresses: Shi Mao, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia, shi.mao@kaust.edu.sa; Yogeshwar Nath Mishra, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia, mishrayn@iitj.ac.in; Wolfgang Heidrich, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia, wolfgang.heidrich@kaust.edu.sa.

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM. This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in *ACM Transactions on Graphics*, <https://doi.org/10.1145/nnnnnnn.nnnnnnn>.

Similar to a liquid tunable lens, DPPs allow for dynamic control of a liquid surface shape for optical purposes. However, instead of only controlling a global lens curvature, DPPs provide fine control over local surface geometry, thereby offering a transmissive alternative to the reflective deformable mirrors used in adaptive optics systems.

In this work we use the ability of the DPP to *dynamically* and *locally* shape an optical wavefront in a compact optical system to demonstrate a new type of foveated imaging system (see Fig. 1). The optical design consists of a combination of a highly aberrated achromatic double lens and a DPP that can be used for aberration correction. This optical system lacks the complexity to *simultaneously* correct for *all* aberrations *everywhere* in the image like a classical camera lens, however, it has the ability to correct for the aberrations in a localized region of interest (the fovea) and produce an image with excellent fidelity in that region. Moreover, through dynamic control of the DPP, this fovea region can be placed anywhere in the image plane using direct electrostatic actuation (c.f. Fig. 1b, Fig. 2). A stack of such foveated images, a.k.a. a *Fovea Stack*, can be used to reconstruct an unaberrated, high fidelity image (c.f. Fig. 1c).

Specifically, we make the following contributions in this work:

- we propose Fovea Stacking as a new paradigm for camera systems; we optimize the DPP's wavefront control patterns for various depths using differentiable optics, resulting in regionally corrected images that can be stacked to produce an aberration-free composite image.
- to efficiently cover the FoV with minimal saccades, we propose a method that jointly optimizes the DPP's deformation patterns, allowing effective stacking of just 3-5 images for full aberration correction.
- to accurately model non-linear DPP behavior for larger control signals, we develop a neural network model for mapping control signals to actual wavefront shapes and utilize this model to reduce the gap between simulation and real experiments.
- we demonstrate extended depth of field imaging by Fovea Stacking across a range of different depths.
- by integrating foveate imaging with object detection or eye tracking, we also demonstrate smooth pursuit movement that dynamically adjusts the imaging system to keep an object of interest within the fovea region, enabling the generation of foveated video suitable for downstream tasks such as surveillance or foveated VR display.
- we demonstrate all the above capabilities on a hardware prototype that is already compact enough to be used outside the lab, although DPPs have not yet reached the same level of miniaturization as liquid tunable lenses.

2 RELATED WORK

2.1 Wavefront Modulators

Wavefront modulators are optical devices capable of dynamically shaping light to compensate for optical distortions. Deformable mirrors (DMs) [Bifano 2011], which are typically reflective, require folded optical paths. Liquid crystal phase-only spatial light modulators (LC SLMs) [Yang et al. 2023], are frequently used for wavefront

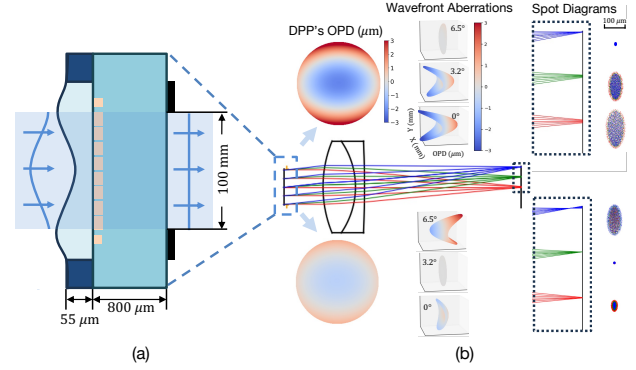


Fig. 2. (a) physical layout of DPP device, composed of a deformable membrane, an optical liquid cavity, and a rigid glass substrate containing hexagonal electrodes. Figure adapted from [Rajaeipour et al. 2021]. (b) localized aberration of different oblique angles can be corrected by dynamically changing DPP patterns. The green-channel wavefront aberration on exit-pupil plane and the corresponding colored spot diagrams are shown.

shaping in microscopy or holography however modern variants are usually reflective designs, since the pixel wiring introduces diffraction artifacts in older transmissive designs. Either variant requires polarized light and exhibits a strong wavelength dependency.

Recently, a novel optofluidic refractive wavefront modulator named a deformable phase plate (DPP) was proposed by [Banerjee et al. 2018]. As shown in Fig. 2a, the DPP consists of a thin optically transparent, liquid-filled cavity sandwiched between a deformable membrane and a rigid glass surface. Its shape can be controlled by exerting electrostatic forces on the membrane via voltages applied to a hexagonal grid of 63 transparent electrodes on the rigid glass back panel. Using this principle, a DPP enables distortion correction up to seventh radial Zernike order [Rajaeipour et al. 2021]. In this work, we adopt the DPP as a refractive wavefront modulator to create a compact imaging system. However, it is observed that traditional linear control strategies are inaccurate for large wavefront deviations. To address this, we propose a neural network-based control strategy to improve accuracy.

Compared to adaptive optics (AO) systems [Hampson et al. 2021; Wang et al. 2018], which typically correct for aberrations introduced by atmospheric turbulence or other external media before light reaches the imaging system, our method focuses on correcting distortions caused by the inherent imperfections of simple, compact imaging optics, tailored to the specific optical system in use.

2.2 Focus Stacking

Focus Stacking extends depth of field by capturing multiple images at varying focus distances. From this image stack, a sharp all-in-focus (AIF) image can be reconstructed. Similarly, Fovea Stacking aims to reconstruct a sharp, aberration corrected image from a stack of regionally sharp images. Focus Stacking can also be used correct for field curvature, where the focal region is a curved manifold instead of a plane. It can therefore serve as a baseline for our work.

Traditional multi-focus image fusion methods reconstruct all-in-focus images by identifying and compositing the sharpest regions

from multiple input images [Huang and Jing 2007; Nayar and Nakagawa 1994]. This fusion can occur in the spatial domain or in transformed domains, such as Laplacian pyramids [Sun et al. 2018; Wang and Chang 2011] and discrete wavelet transforms [Pradnya and Sachin 2013]. For comprehensive reviews of traditional and deep learning-based methods, see [Zafar et al. 2020; Zhang 2021]. Alternatively, some approaches jointly reconstruct both a depth map and an all-in-focus image - for example, Wang et al. [2021] use 3D convolution to learn attention weights from the focus stack, which are then used to produce both the AIF image and the corresponding depth map.

Traditional focus adjustment often involves mechanical movement of the sensor or lens elements, which can induce vibrations during large-amplitude, high-frequency motion [Kuthirummal et al. 2010]. As an alternative, focus-tunable liquid lens (FTLs) have been adopted for focus-sweeping applications [Miau et al. 2013], but they are limited to tuning optical power without the ability to correct higher-order aberrations. Although the adjustable range of the DPP is limited, it is capable of adjusting the focus distance by modulating the "defocus" term in the Zernike polynomial. Our method, therefore, supports stacking not only in the x-y direction but also along the depth direction, enabling aberration-corrected imaging for extended depth-of-field applications.

2.3 Differentiable Optics

Recent advances in computational optics demonstrate the effectiveness of end-to-end co-design of optics and reconstruction algorithms [Sitzmann et al. 2018; Sun et al. 2021]. Central to this approach is a differentiable optics model. The first differentiable implementation of light propagation through a diffractive optical element (DOE) [Sitzmann et al. 2018] enabled applications in extended depth of field and super-resolution. This DOE model has since supported optical designs for wide field-of-view imaging [Peng et al. 2019], coded super-resolution SPAD imaging [Sun et al. 2020], and multi-aperture hyperspectral imaging [Shi et al. 2024]. To support the end-to-end design of compound refractive lenses beyond simple wave-optics systems, researchers introduced both neural proxies [Tseng et al. 2021] and differentiable ray-tracing [Sun et al. 2021], which computes ray-surface intersections and refractions with full differentiability with respect to lens parameters. Subsequent work on differentiable ray-tracing reduced memory usage [Wang et al. 2022] and enabled automatic lens design from random initialization [Yang et al. 2024a], leading to the design of compound fluidic-freeform lenses [Na et al. 2024] and even hybrid refractive-diffractive lenses [Yang et al. 2024b].

In this work, we adopt a differentiable ray-tracing framework [Yang et al. 2024a] to optimize DPP deformation patterns for foveated imaging and stacking, targeting reconstruction quality. The customized differentiable model for the DPP device is detailed in Sec. 3.

2.4 Foveated Imaging

Foveated imaging, inspired by human vision, concentrates high resolution where it is most needed - similar to the function of the fovea. Many different approaches have been proposed to achieve spatially varying resolution through specialized sensor design or image

fusion techniques. Retina-like sensors, for example, arrange photoreceptors in a spatially varying log-polar structure [Boluda et al. 1996; Sandini and Metta 2003]. Imaging systems using prism arrays [Carles et al. 2016] or a Risley-prism [Huang et al. 2021] expand the FoV by deflecting incoming rays and digitally super-resolving the overlapping central region, effectively producing foveated images through fusion. Similarly, a 3D-printed microlens system [Thiele et al. 2017] combines pixels from four microlenses with different focal lengths to generate a foveated image. However, these methods hard-code the fovea region, necessitating mechanical camera motion to perform saccades.

More directly related to our work is a small body of literature on creating foveated imaging systems by locally correcting optical aberrations using dynamic wavefront shaping methods [Martinez et al. 2001]. Such foveated images result from localized aberration correction, which can effectively address only one limited region of interest at a time. To correct off-axis aberrations locally, various optical solutions using different wavefront modulators have been proposed, including reflective SLMs [Curatu et al. 2005; Wick et al. 2002], transmissive SLMs [Harriman et al. 2006], and reflective DMs [Zhao et al. 2008a,b]. However, these methods are either bandwidth-limited and require polarized light (as with SLMs), or they require folded optical paths and complex system designs (as with reflective modulators), which is counterproductive for miniaturizing and simplifying optical systems. Furthermore, they typically optimize the optics for focusing at infinity and at a fixed oblique angle. In this work, we propose the use of a refractive wavefront modulator to enable compact and wide-band imaging. Additionally, we introduce a differential optics-based optimization method to tailor the desired wavefront across different imaging distances.

3 WAVEFRONT OPTIMIZATION WITH DIFFERENTIABLE OPTICS

In this section, we describe how to optimize DPP wavefront control patterns to (1) dynamically position the fovea anywhere within the image and (2) efficiently cover the field of view (FoV) with minimal saccades. We develop a forward model of the optical system shown in Fig. 1, comprising an achromatic doublet lens and a deformable phase plate, to enable ray tracing through the system. This model is implemented in a differentiable manner, allowing us to optimize the DPP's wavefront parameters via back-propagation of the imaging loss. The subsections are organized as follows: Sec. 3.1 details the DPP model within differentiable optics system; Sec. 3.2 covers the optimization of a single DPP wavefront for foveated imaging at oblique angles. Sec. 3.3 presents strategies for jointly optimizing multiple DPP wavefronts to cover the FoV for Fovea Stacking, under a limited number of images.

3.1 DPP Surface Modeling

With an optofluidic cavity thickness of $55\mu\text{m}$, the DPP can be treated as a thin plate capable of phase modulation. Based on Fermat's principle, the generalized Snell's law of refraction [Yu et al. 2011] governs the direction of light propagation in response to such abrupt phase changes. Accordingly, in the differentiable optics system, we

model the DPP as a refractive plane that bends light according to the optical path difference (OPD) it introduces.

$$\begin{aligned}\alpha' &= \alpha + \frac{\partial D(x, y)}{\partial x} \\ \beta' &= \beta + \frac{\partial D(x, y)}{\partial y} \\ \gamma' &= \sqrt{1 - \alpha'^2 - \beta'^2}\end{aligned}$$

Here, $\mathbf{n}' = (\alpha', \beta', \gamma')^T$ and $\mathbf{n} = (\alpha, \beta, \gamma)^T$ are the normalized directions of in-going and refracted rays, respectively, and D denotes OPD. To represent the wavefront generated by the DPP, we parameterize the OPD as Zernike polynomials up to K polynomials:

$$D(\rho, \varphi) = \sum_k^K w_k Z_k(\rho, \varphi), \quad (1)$$

where ρ is normalized radial distance (normalized by aperture radius), φ is the azimuthal angle, Z_k and w_k are the k -th Zernike polynomials and corresponding coefficients following OSA indices [Thibos et al. 2002]. Partial derivatives with respect to x and y can be therefore derived and parameterized by w_k . Because horizontal and vertical tilt only cause geometric distortion – resulting in unwanted pixel misalignment across images rather than improved image sharpness – they are excluded from Zernike coefficient optimization and fixed at zero. Notice that we do not correct for radial lens distortion since it remains consistent across a stack of images as long as the DPP does not introduce tilt. More details on Zernike polynomials and their derivative calculation is provided in the Supplement.

3.2 Optimized Wavefronts for Foveated Imaging

To maximize the imaging quality at a specific oblique angle, corresponding rays are sampled and traced through the optical system, consisting of a DPP and a lens. The RMS spot size serves as the optimization loss, guiding the optimization of the DPP wavefront pattern via backpropagation. Specifically, for each sampled object point P at distance D , N_a points $\{P_i | 1 \leq i \leq N_a\}$ are sampled on the aperture plane to generate the sampling rays $\{R_i = \text{Ray}(P \rightarrow P_i) | 1 \leq i \leq N_a\}$. Each sampled ray is then traced through the optical stack to reach the image plane at positions $p_{i\lambda} = RT(R_i; \lambda)$, given a specific wavelength λ . The root-mean-square radial deviation of traced rays from their center, termed the **RMS spot size**, is used as the loss function minimized via backpropagation. As RMS spot size is very closely related to the area under the modulation transfer function MTF [Luc 2022], it is a popular loss in lens design. For color images, the RMS spot size is further integrated over multiple wavelengths.

$$r(P) = \frac{1}{3} \sum_{\lambda} \frac{1}{N_a} \sum_i^{N_a} \|p_{i\lambda} - \bar{p}\|_2 \quad (2)$$

To enhance convergence, we reduce the learning rate for each Zernike coefficient by a factor of $\sqrt{10}$ for each successive polynomial order. Since the horizontal/vertical tilting terms only shift the image and do not introduce optical degradation, they are excluded to prevent ray deflection.

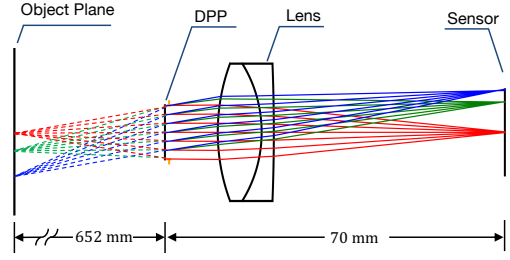


Fig. 3. Side view of the calibrated simulation system. The object plane is set 652 mm away. The imaging setup includes a DPP, a 50 mm achromatic doublet lens (Thorlabs AC-254-050A), and a Bayer pattern RGB sensor (FLIR GS3-U3-41C6C-C).

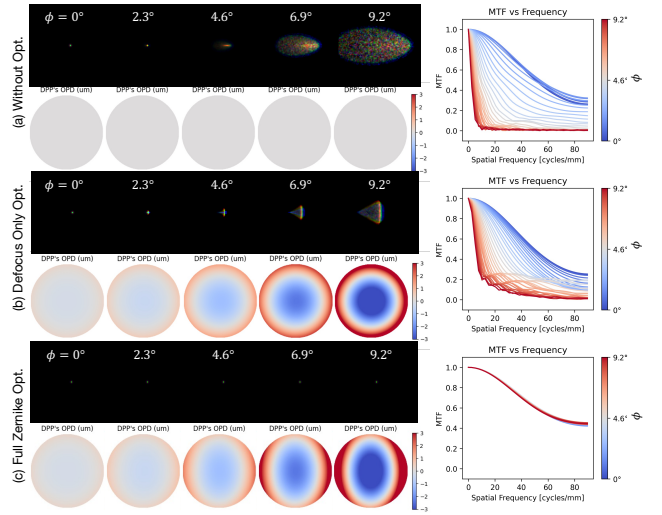


Fig. 4. Optimized PSF and MTF for various oblique angles ϕ relative to the optical axis: (a) initial system without DPP, (b) optimization of the defocus Zernike parameter only, simulating improvements that can be made by focus adjustments (either by shifting the lens or with a liquid tunable lens) and (c) full Zernike optimization up to 4th order. MTF values are averaged over sagittal and tangential components.

To verify the DPP's capability for localized aberration correction, an optical system is set up and calibrated in a real-world setting as described in Sec. 6, and shown in Fig 3. A total of 32 oblique angles ϕ are sampled within the range of 9.2° (corresponding to half the sensor's diagonal size R) along the horizontal axis. The PSFs for five angles of these angles are visualized in Fig 4. The average modulation transfer function (MTF) is computed from the PSFs by first performing a weighted average of the RGB channels based on relative luminance, followed by averaging the sagittal and tangential MTFs. The initial setting without any optimization, exhibits severe off-axis aberrations (Fig 4a). While adjusting the focus - equivalent to optimizing only the defocus term in Zernike polynomials - can mitigated field curvature (Fig 4b), other aberrations remain uncorrected. Full optimization up to the 4th order of Zernike polynomials effectively corrects these off-axis aberrations (Fig 4c).

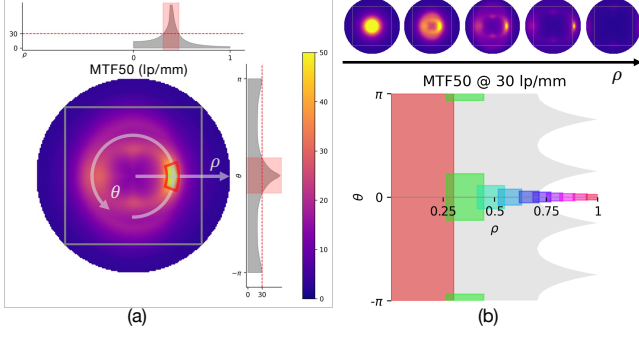


Fig. 5. Radial and angular coverage of the fovea area. (a) For a single optimized DPP pattern, the fovea area—defined as the area where MTF50 exceeds a threshold like 30 lp/mm—can be characterized by its extent along the normalized radial axis ρ and angular axis θ . The gray square indicates the sensor region. (b) As the optimized angle approaches the sensor's edge, the fovea area shrinks, necessitating more images in the peripheral areas to maintain sharpness over the gray area representing the sensor.

Although aberration can be effectively corrected locally for different oblique angles ϕ , the size of the fovea varies by location. In Fig 5, for each optimized DPP wavefront pattern, the MTF50 (defined as the spatial frequency at which the MTF drops to 50%) is visualized within a circular FoV with radius R , which circumscribes the square sensor (shown as a gray square). Given a specific sharpness threshold (e.g., 30 line pairs per mm), the radial and angular regions exceeding this threshold are identified and plotted as a rectangular region in a normalized radial-angular diagram. The normalized radius ρ is related to the angle ϕ by: $\rho = f \tan(\phi)/R$, where f is the effective focal length. The sharp coverage area tends to shrink as it approaches the periphery. It is estimated that approximately 100 images are needed to cover the entire square sensor region (shaded in gray in the radial-angular diagram) with resolution exceeding 30 lp/mm. More detailed analysis is provided in Supplementary. To capture sharp images across the full FoV more efficiently, a trade-off between sharpness and coverage must be considered. Notably, a secondary peak in sharpness often appears in the opposite direction (as shown at $\theta = \pm\pi$ in Fig. 5a), due to the known symmetries of certain aberrations, which can potentially be exploited.

3.3 Optimized Wavefronts for Fovea Stacking

3.3.1 ROI tiling. Aberration correction across the field of view can be achieved by dividing the FoV into regions of interest (ROI) and correcting local aberrations independently. To optimize each wavefront for local correction within a ROI at distance D , M points are sampled - typically on a grid - within the ROI. The optimization loss, L_{ROI} , is defined as the mean RMS spot size across these sampled points:

$$L_{ROI} = \frac{1}{M} \sum_m r(P_m) \quad (3)$$

3.3.2 Joint optimization. Although full-aberration correction can be achieved by tiling a grid of independently corrected ROIs, this approach does not efficiently exploit the information available across the entire image stack. Therefore, to jointly optimize aberration correction across the full FoV with a limited image budget N , we propose to optimize the RMS spot size grid across the stack collectively.

We first sample a grid of $H \times W$ points over the entire FoV at a given depth. Ray-tracing is then performed for each of the N different DPP wavefronts, and the corresponding RMS spot size is calculated, resulting in a stack of RMS spot size grid $\{r_n(i, j) \in R^{H \times W} | 1 \leq n \leq N\}$. Each element $r_n(i, j) = r_n(P_{ij})$ represents the RMS spot size of the sampled point P_{ij} , computed using the n -th DPP wavefront. The overall image quality after fusion depends on the minimal RMS spot size at each location across the stack, i.e. $r_{min}(i, j) = \min_n r_n(i, j)$. Therefore, to jointly optimize for all DPP patterns, we minimize the **grid stacking** loss.

$$L_{gs} = \frac{1}{H \times W} \sum_{i,j} r_{min}(i, j) \quad (4)$$

Essentially, the grid stacking loss ensures that for each wavefront, only the area where it performs best across the entire stack will be considered as its active region for optimization. However, optimization may yield some degenerate patterns where does not outperform all other waveforms anywhere in the image, thus having no active region at all. This usually happens when the budget of images increases. To alleviate it, we encourage these DPPs to optimize on the **hard regions**, using r_{min}/\bar{r}_{min} as a weight:

$$L_{hr} = \sum_{n'} \frac{1}{H \times W} \sum_{i,j} r_{n'}(i, j) \frac{r_{min}(i, j)}{\bar{r}_{min}} \quad (5)$$

$$n' \in \{n | \sum_{i,j} \mathbb{1}(n, n^*) = 0\},$$

where $\mathbb{1}(\cdot)$ is the identity function, and $n^*(i, j) = \operatorname{argmin}_n r_n(i, j)$ is the mask index for the sampled point, indicating which DPP wavefront provides the best image quality for this location. The overall loss for **joint optimization** is a sum of the two losses:

$$L_{joint} = L_{gs} + L_{hr} \quad (6)$$

The joint optimization scheme is validated using the same simulated optical system, as shown in Fig. 6. In the experiment, 32×32 grids are sampled across the full FoV to jointly optimize five wavefront patterns. The resulting optimized PSF grids exhibit compensatory patterns due to the gradients being masked differently by the index n^* to optimize each DPP wavefront. As expected, the PSF grid displays radial symmetry, which the dynamic optimization scheme efficiently exploits to cover the entire FoV within a limited imaging budget.

3.3.3 Efficiency under limited imaging budget. The efficiency of using the imaging budget is analyzed in Fig 7. With a limited image count, we optimize DPP patterns and evaluate overall imaging quality using the average RMS spot size of the global PSF across the stack—i.e., the mean r_{min} (blue). ROI tiling requires selecting non-overlapping ROIs to cover the entire FoV; thus, image budgets

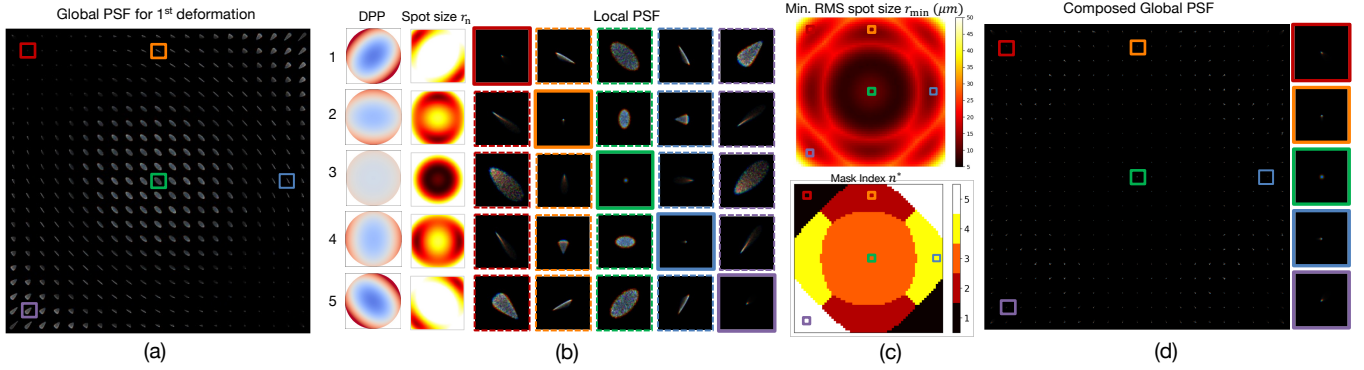


Fig. 6. Joint optimization of 5 DPP patterns using grid stacking. (b) After optimization, each DPP deformation is specialized to optimize quality over different local regions while jointly covering the full FoV. (a) As an example the 1st DPP pattern produces the best focus in the top-left and bottom-right regions. (c) The minimal value of RMS spot size r_{min} across the stack serves as the optimization loss, while the mask index n^* indicates which DPP pattern is best for any given region, which can be used to generate (d) a composite PSF map for the image after stacking.

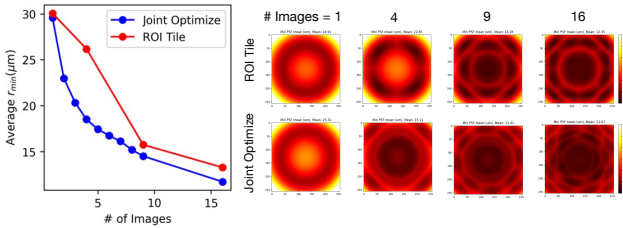


Fig. 7. Comparison of the imaging efficiency between joint optimization and ROI tiling.

of 2×2 , 3×3 , and 4×4 are included for comparison (red). Joint optimization significantly improves image quality with as few as 3-5 images, whereas ROI tiling proves less efficient. Beyond 10 images, further improvements in image quality become marginal.

4 DPP CONTROL MODEL

To control the DPP's deformation, earlier work used a linear model [Banerjee et al. 2018; Rajaeipour et al. 2021], which assumes the desired membrane shape is related to the square of the electrode voltages via a linear influence matrix. i.e.

$$W = AV^2, \quad (7)$$

where $W \in R^K$ represents the Zernike coefficients, and $V^2 \in R^{63}$ is the element-wise square of the electrode voltages. The matrix $A \in R^{K \times 63}$ is the influence matrix obtained through calibration. To calculate the voltage corresponding to a desired set of Zernike coefficients W , it is therefore modeled by solving a constrained optimization problem such that the voltage is within the physical limit $[0, V_{max}]$:

$$\min \|W - AV^2\|_2^2 \quad s.t. 0 < V < V_{max}, \quad (8)$$

where $V_{max} = 270V$ is the maximum voltage for each electrode. This model is also implemented in the drivers of the DPP.

However, this model has two downsides: first, since the hardware capabilities limited by V_{max} , working in Zernike space makes it

difficult to predict whether a desired shape is feasible on the device. Second, even if the target shape within the feasible range, we find that the linear model breaks down for larger amplitudes. To address this limitation, we propose a neural network-based control strategy that captures the non-linear behavior of the device and improve accuracy, thereby reducing the gap between physical implementation and simulation. Furthermore, by controlling the shape directly in voltage space, the feasibility of the desired wavefront correction is guaranteed.

In the following subsections, we propose our neural networks for DPP Control in Sec. 4.1, then the DPP's wavefront is measured using the experiment setup described in Sec. 4.2 for neural network training in Sec. 4.3, finally the control strategies is analyzed in Sec. 4.4.

4.1 Neural Networks for DPP Control

To calibrate the wavefront response with respect to the control voltage applied to the device, we experimentally apply different voltages and measure the resulting optical path differences using a customized wavefront sensor [Wang et al. 2019]. Given a known aperture size, Zernike coefficients are fitted to reconstruct the measured wavefronts. By collecting a dataset of paired voltage inputs and corresponding Zernike coefficients, we train encoder-decoder neural networks to control the DPP.

The decoder predicts wavefront coefficients from control voltages using a dual-branch architecture: a linear branch to model overall linearity and a sigmoid-activated branch to capture residual nonlinearity. This residual learning strategy builds upon the linear model's approximate estimation [Rajaeipour et al. 2021], requiring only nonlinear refinements. The input is extended to $[V, V^2]$, to facilitate pattern recognition. Conversely, the encoder, which predicts the desired control voltages from wavefront coefficients, is modeled as a non-linear mapping using a simple two-layer multilayer perceptron with a final sigmoid activation to constrain the output within the physical voltage limits. Voltage normalization by V_{max} ($\tilde{V} = V/V_{max}$) is applied to ensures stable training.

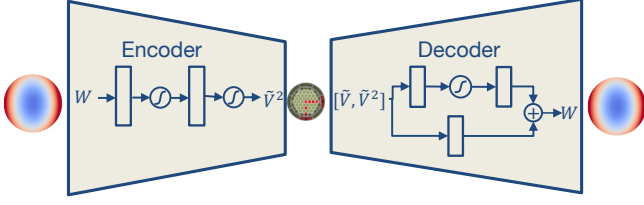


Fig. 8. Neural networks for DPP control. Rectangular blocks stand for linear layers, activation functions are sigmoid functions. The decoder's residual design enables learning non-linear refinements based on the linear model's approximate estimation.

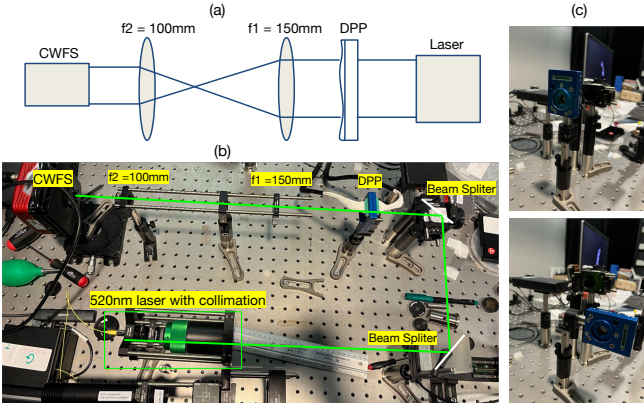


Fig. 9. Wavefront Measurement Setup. (a) The optical path of the Measurement system. A $4f$ system with a spatial magnification of $-2/3$ relays the DPP device's aperture plane to the wavefront sensor. (b) The realworld relays the DPP's aperture to the CWFS, with alignment ensured by an auto-collimator. Lenses with focal lengths of 150 mm and 100 mm provide a spatial magnification of $-2/3$, effectively projecting the full 10 mm DPP aperture onto most of the CWFS sensor area (12.46×6.7 mm). Illumination is provided by a collimated 520 nm laser source. The DPP is mounted on a flip-mount that can rotate 90° , allowing it to be removed from the optical path to provide a collimated wavefront reference. The CWFS then measures the DPP's relative wavefront changes. From the measured relative wavefront, we fit the Zernike polynomials within the circular aperture region to get the corresponding coefficients.

4.2 Wavefront Measurement

A coded wavefront sensor (CWFS) [Wang et al. 2019] is used to measure wavefront changes in the DPP. As shown in Fig 9, a $4f$ system relays the DPP's aperture to the CWFS, with alignment ensured by an auto-collimator. Lenses with focal lengths of 150 mm and 100 mm provide a spatial magnification of $-2/3$, effectively projecting the full 10 mm DPP aperture onto most of the CWFS sensor area (12.46×6.7 mm). Illumination is provided by a collimated 520 nm laser source. The DPP is mounted on a flip-mount that can rotate 90° , allowing it to be removed from the optical path to provide a collimated wavefront reference. The CWFS then measures the DPP's relative wavefront changes. From the measured relative wavefront, we fit the Zernike polynomials within the circular aperture region to get the corresponding coefficients.

4.3 Training of DPP Control Models

We captured 1800 voltage-Zernike coefficient pairs by randomly sampling Zernike coefficients within the device's operating range. We used 80% of the data for training and the remaining 20% for testing. More details on dataset collection are provided in the supplement. For comparison, a linear decoding model was trained alongside our neural-network (NN) model. To address measurement error

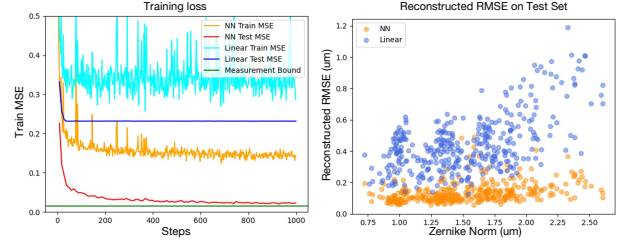


Fig. 10. Decoder training loss and reconstruction error on the test set. (a) Training loss for the NN model is significant lower than for the linear model and approaches the measurement bound. (b) The NN model maintains lower reconstruction error for the full range of deformations, while the linear model tends to performs worse for larger deformations (i.e. of larger Zernike norm).

and prevent overfitting, random noise with a standard deviation of 0.36 was added to the Zernike coefficients for data augmentation. As illustrated in Fig 10, the neural network model achieves significantly lower training and testing loss than the linear model, approaching the measurement limit established by repeated CWFS measurements. On the test set, the linear model's reconstruction error (RMSE) increases with larger amplitudes, indicating poor performance at higher magnitudes, whereas the neural network maintains a more consistent error across magnitudes.

Unlike the linear model, which uses constrained least squares to compute the control voltage from the desired Zernike coefficients, we train an encoder neural network to predict the control voltage directly. This encoder achieves an RMS error of 2.56 on the test set (in voltage control units).

4.4 Control Strategies

We compare different strategies for controlling the DPP to obtain the control voltage corresponding to a desired target Zernike Coefficient W_t , including the linear control strategy as a baseline. As shown in Fig. 11, the "encoder" strategy directly predicts the control voltage V via encoder neural network inference. The "decoder" strategy initializes V as a zero tensor and then optimize it through back-propagation using the decoder, with W_t as the target. The "encoder+decoder" strategy initializes V as the predicted value from encoder and perform the following back-propagation through the decoder. For 10 known voltage-Zernike coefficient pairs in the test set, we applied each control strategy to obtain corresponding control voltages V . These voltages were then applied to the DPP device, and the resulting wavefronts V_m were measured as described in Sec. 4.2.

Table 1 presents the mean square error for various control strategies. The reconstructed voltage V and model-predicted Zernike wavefront W_r were evaluated across the entire test set (360 samples), while V_m was measured for 10 target Zernike wavefronts. The encoder-decoder approach achieves the lowest control voltage error and consequently yields the most accurate wavefront measurement (W_m), particularly minimizing wavefront errors for higher-order Zernike aberrations (2nd order and above, excluding tilt). Figure 11 provides a qualitative example: the 63 control voltages are shown

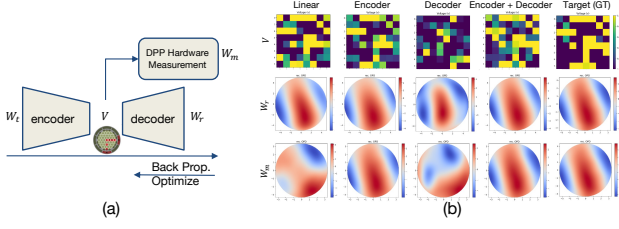


Fig. 11. Comparison of different control strategies. (a) Given a target wavefront W_t , the control voltage V is acquired by different control strategies (see text), and a prediction of the reconstructed wavefront W_r is obtained from the model. To verify the effectiveness, the control signal V is sent to the DPP device to obtain the measured wavefront W_m . (b) An example of different control strategies.

MSE	Linear	Encoder	Decoder	Enc. + Dec.
V	3.280	2.560	17.440	2.140
W_r	0.106	0.119	0.217	0.002
W_m	5.667	0.166	1.160	0.103
W_m (h. o.)	5.187	0.099	1.098	0.013

Table 1. Quantative comparison of different control strategy. h.o. denotes high orders, meaning second order and above.

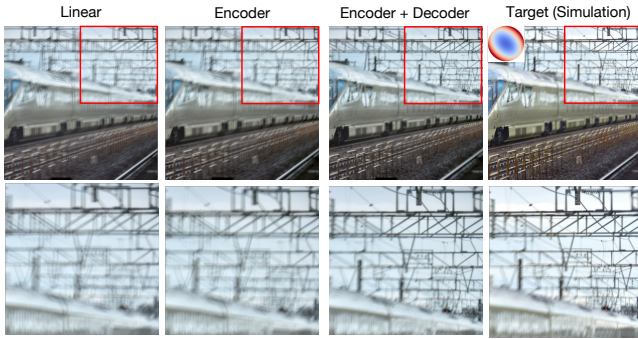


Fig. 12. The captured images using different DPP control strategies. The encoder-decoder strategy best aligns with the target in simulation.

in an 8×8 grid, with the last entry set to zero. Note that this grid is a visualization and does not represent the actual physical layout of the electrodes.

We further verified the control strategies using the calibrated simulation model. As shown in Fig 12, the target image is rendered in simulation using the desired wavefront pattern. Among all control strategies, the encoder+decoder method produces results that best align with the target image.

5 IMAGE STACKING

To fuse the captured image stack, a straightforward approach involves pixel-wise fusion using the mask index n^* obtained from the joint optimization process. However, real-world hardware imperfections - such as deviations in wavefront fidelity and calibration error - can lead to discrepancies between the simulated and actual

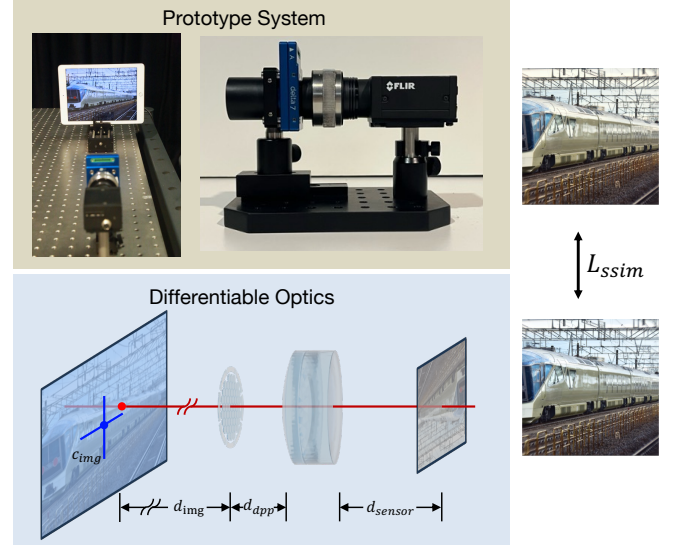


Fig. 13. Prototype system and calibration. Our system comprises a DPP, an achromatic lens within a zoom enclosure for adjustable working distances, and a sensor. Assembly distances were calibrated by matching captured and simulated images.

system, degrading image quality when using the pre-optimized mask alone. Additionally, for extended depth-of-field applications, scene-dependent depth variations necessitate a flexible stacking algorithm.

While Fovea Stacking shares conceptual similarities with Focus Stacking, deep learning-based Focus Stacking methods are often trained on spatial-invariant, disk-pattern blur kernels, which offer limited improvement over traditional sharpness-based techniques. As a result, we adopt a conventional image fusion pipeline. We identify sharp regions by applying a Laplacian filter, followed by blurring. Each pixel in the fused image is computed as a weighted average across the image stack, with weights determined by sharpness. Further implementation details are provided in the supplement. Experiments demonstrate that, despite its simplicity, this stacking method is effective in producing high-quality, aberration-corrected images.

6 PROTOTYPE SYSTEM AND CALIBRATION

Our prototype comprises a DPP (Phaseform Delta 7), a doublet lens, and an image sensor. As shown in Fig. 13, the lens is housed in a zooming enclosure (Thorlab SM1ZM) to adjust the back focus distance (d_{sensor}) for different working distances. A 50 mm achromatic doublet lens (Thorlab AC-254-050-A) was selected for its simplicity and inherent achromatic aberration compensation. The Bayer-pattern RGB sensor (FLIR GS3-U3-41C6C-C) features a pixel size of $5.5 \mu\text{m} \times 5.5 \mu\text{m}$ with 2048×2048 resolution. The total length of the compact imaging system is 15 cm. Since the components are assembled using threaded lens tubes, the relative distances between optical elements require further calibration.

We calibrate the prototype system using an image-based method that leverages differentiable optics to optimize calibration parameters by matching captured images with simulated counterparts. This process addresses unmeasured distances within the camera setup: the distance between the DPP and the lens's front plane (d_{dpp}), and the distance between the lens's back plane and the sensor (d_{sensor}). The imaging model also incorporates the distance between the image plane and the DPP (d_{img}) and the shift between the image center and the intersection of the object plane with the optical axis (c_{img}). Given known screen and sensor dimensions, these parameters are jointly optimized. To ensure robust calibration, we capture images using multiple DPP patterns and employ a translation stage to precisely move the image plane. Due to differences in display characteristics and sensor response, the pixel values of captured and simulated images may differ. Therefore, we use the structural similarity index measure (SSIM) to compare local image structure rather than pixel values. All distances are initialized using rough physical measurements, and the image shift c_{img} is calculated from a homography transform estimated by matching the captured image to the known screen pattern.

7 EXPERIMENTS

In this section, we experimentally demonstrate the capabilities of the hardware prototype in three applications: aberration-corrected imaging via Fovea Stacking at far distances (Sec. 7.1), extended depth-of-field imaging via Fovea Stacking across various depths (Sec. 7.2), and foveated object tracking during smooth pursuit movements (Sec. 7.3). Throughout the experiments, all images were captured in RAW and demosaiced using bilinear interpolation. We did not correct lens distortion because it remains consistent across images, and the DPP tilting terms were set to zero to prevent pixel shifts.

7.1 Fovea Stacking

For imaging beyond the hyperfocal distance, the sensor plane is re-positioned via zoom housing rotation. Five phase patterns are then jointly optimized at distance 60m as described in Sec. 3.3 and used for image capture. For comparison, a focus stack with the same number of images was optimized using only defocus as a variable Zernike polynomial term. As shown in Fig. 14, single images are affected by significant off-axis aberrations (Fig. 14a). Although Focus Stacking can partially compensate for field curvature, it does not fully correct these aberrations (Fig. 14b). Fovea Stacking offers better aberration correction, as demonstrated by substantially improved resolution in fine structures and enhanced text readability in this comparison (Fig. 14c). The stacking use the sharpness-based method described in Sec. 5.

Fig. 15 compares sharpness-based fusion (3rd column) with fusion using a pre-optimized mask (last column), which directly selects pixel values from the pre-optimized mask as described in Sec. 3.3. Real-world hardware imperfections introduce minor artifacts at image boundaries (most apparent in the words "the," "increasing," and "reducing"). In contrast, sharpness-based fusion mitigates these artifacts by providing smoother transitions at the boundaries. The sharpness map for each image aligns well with its pre-optimized mask, making it a reliable indicator for fusion.

7.2 Fovea Stacking for Extended Depth of Field

To achieve extended depth of field using Fovea Stacking (Fig. 16), we sampled three planes evenly in disparity space between 535 and 835 mm, optimizing five phases per plane for a total of 15 images. Four objects were placed at varying depths, with two car models positioned at angles to enhance the depth variation. For a fair comparison with the Focus Stacking method, we also optimized 15 phases, varying only the defocus term in the Zernike polynomials.

Fig. 17 shows that a single capture (Fig. 17a) focused around 652 mm suffers from significant off-axis and out-of-focus aberrations. While Focus Stacking (Fig. 17b) mitigates out-of-focus aberrations, off-axis aberrations still limit resolution in peripheral areas. Fovea Stacking (Fig. 17c) corrects for both aberrations, and preserves finer details.

For extended depth of field application, fusion with pre-optimized masks is unsuitable, as object depths vary arbitrarily. In Fig. 18, we compare our fusion method to Laplacian Pyramid [Wang and Chang 2011], IFCNN [Zhang et al. 2020], MGDN [Guan et al. 2023], and DERE [Si et al. 2023]. Pairwise fusion networks such as IFCNN and MGDN lose sharpness when sequentially fusing the entire image stack. Although DERE processes the whole stack simultaneously, we observed that it overfits its training data and produces hue variations depending on the focus depth of each image. While the traditional Laplacian Pyramid method achieves comparable sharpness, it is less effective in text readability and barcode resolution. We attribute the lower quality of neural network-based methods to the introduction of different types of PSF in fovea imaging, compared to their original Focus Stacking domain where optics are typically well corrected for off-axis aberrations. In contrast, traditional sharpness-based and Laplacian-based methods are less sensitive to the blur kernel, resulting in better fusion quality.

7.3 Foveated Imaging for Object Tracking

For surveillance video applications, where only a single object in the field of view is of interest. Our method tracks this object and ensures its corresponding area is imaged clearly. In our experiment, a pink car moves at 1 mm/s on a translation stage, positioned 652 mm from the camera. After selecting the object region in the first frame, we employ a tracking algorithm [Bolme et al. 2010] to follow the target's bounding box throughout the sequence. For tracking application, the control voltage that locally corrects aberrations must be computed quickly to minimize latency. Because back-propagation optimization at each location is too slow for real-time use, we instead pre-optimize DPP deformation patterns on a 9×9 grid across the image and use linear interpolation to estimate the control voltage at any arbitrary location. Further analysis of the accuracy of interpolation control is provided in the supplement. For fovea tracking, the interpolated control voltage at the bounding box center is applied to the DPP device for image capture. For focus tracking, voltage is interpolated similarly, but using grids optimized only for defocus correction. Although DPP device have a response time of less than 55 ms, we wait for 200 ms before capturing the next frame to ensure stabilized deformation.

As shown in Fig. 19, without DPP modulation, the optical system with a focus distance at 652 mm is blurred in off-axis regions

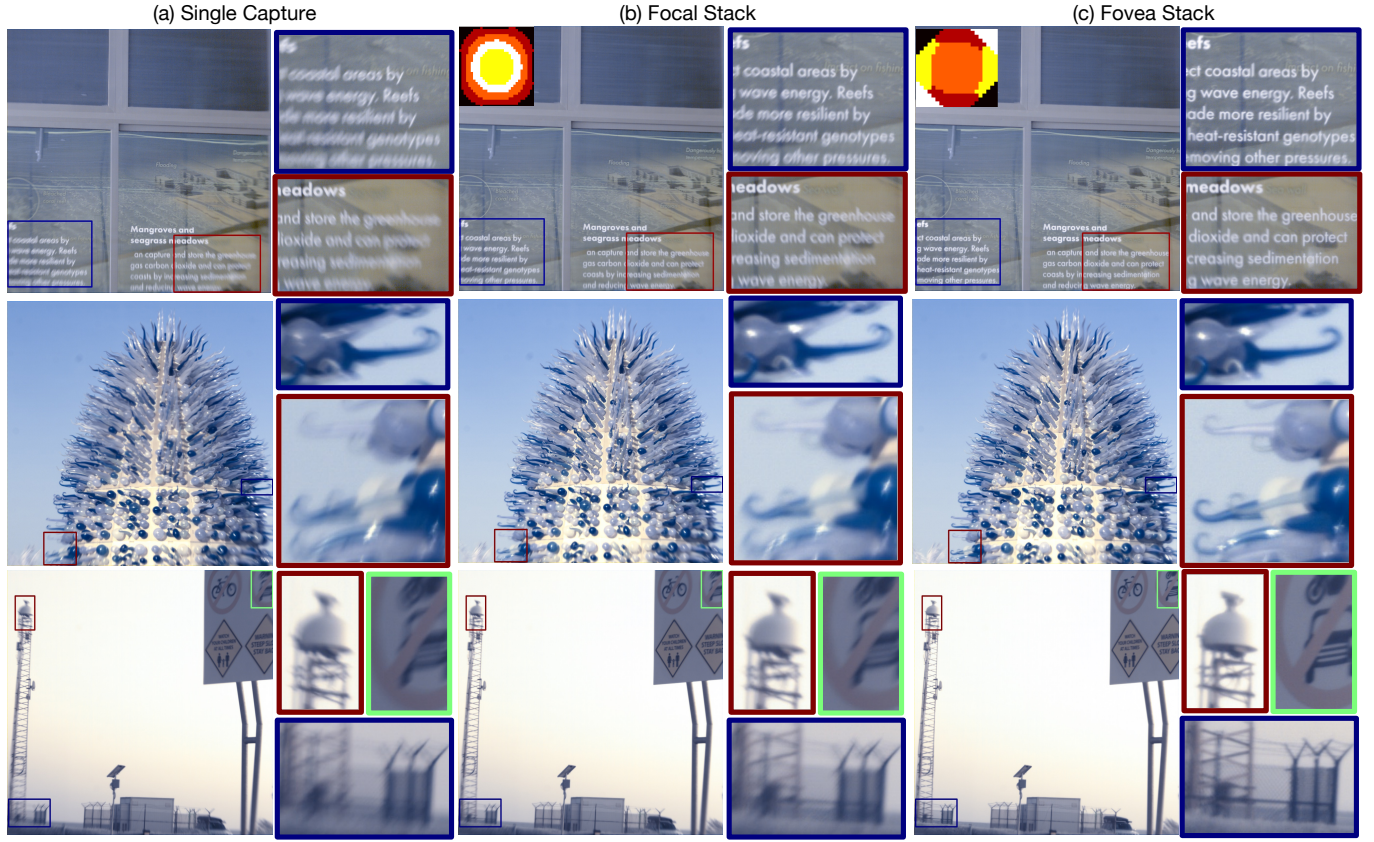


Fig. 14. Fovea Stacking: (a) A single capture suffer from significant off-axis aberration. (b) Focus Stacking 5 images partially mitigates the curvature of field issue and thus improve the imaging quality, but doesn't eliminate the aberrations. (c) Fovea Stacking provides superior aberration correction. Mask index is visualized for stacking.

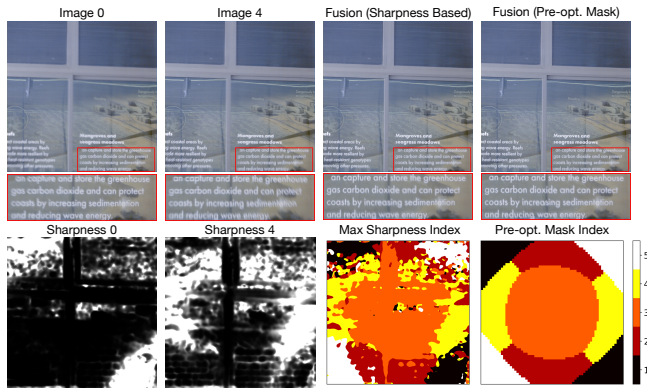


Fig. 15. Comparison of Fusion by Image Sharpness vs. Pre-optimized Mask: The first two columns show the captured image I_k and its corresponding sharpness map S_k . The third column presents the sharpness-based fusion result, along with the max sharpness index for each pixel. The final column displays fusion using a pre-optimized mask, which introduces noticeable artifacts along region boundaries. Artifacts are most apparent in the words "the," "increasing," and "reducing."

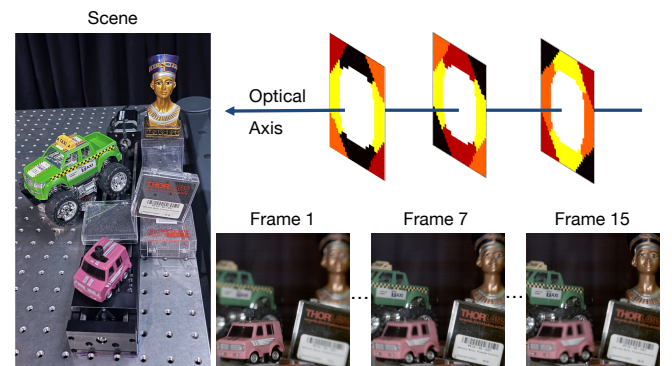


Fig. 16. The Scene and scheme for extended depth of field. The scene composes 4 objects distributed within the range of 535 to 835 mm. 15 DPP patterns are optimized for three distances, each have 5 different patterns.

(Fig. 19a). Adaptive focus tracking corrects field curvature, improving local image quality (Fig. 19b). Fovea Stacking maintains sharpness throughout the video (Fig. 19c). Please also see the accompanying video.



Fig. 17. Fovea Stacking for Extended Depth of Field. (a) A single capture is limited by both out-of-focus blur and off-axis aberrations. (b) Focus Stacking reduces out-of-focus blur but does not correct off-axis aberrations. (c) Fovea Stacking corrects both, as seen in the improved sharpness and text clarity.



Fig. 18. Comparison of different fusion methods for extended depth of field application. Pairwise fusion networks (IFCNN, MGDN) suffer from quality degradation due to sequential fusion. DEReD, although processing the entire stack, is prone to overfitting. Traditional Laplacian Pyramids offer improved sharpness, but still struggles in text readability and barcode resolution. Our method achieves the highest overall quality.

8 CONCLUSION

We have presented a novel imaging approach that leverages a deformable phase plate (DPP) for dynamic, local aberration correction within a compact optical system. Our method introduces Fovea Stacking as a new paradigm for camera systems: by optimizing DPP wavefront control patterns using a differentiable optical model, we produce regionally corrected images across varying depths that can be stacked to form a high-quality composite image with drastically reduced aberrations.

To efficiently cover the field of view with minimal saccadic movements, we proposed a joint optimization framework for DPP deformation patterns, enabling full aberration correction from as few as 3–5 stacked images. To address the non-linear behavior of DPPs, particularly for larger control signals, we developed a neural network-based control model that maps desired wavefronts to actuation patterns, thereby bridging the gap between simulation and real-world performance.

Our experiments demonstrate the system’s capabilities in aberration-corrected imaging and extended depth-of-field imaging. Comparing

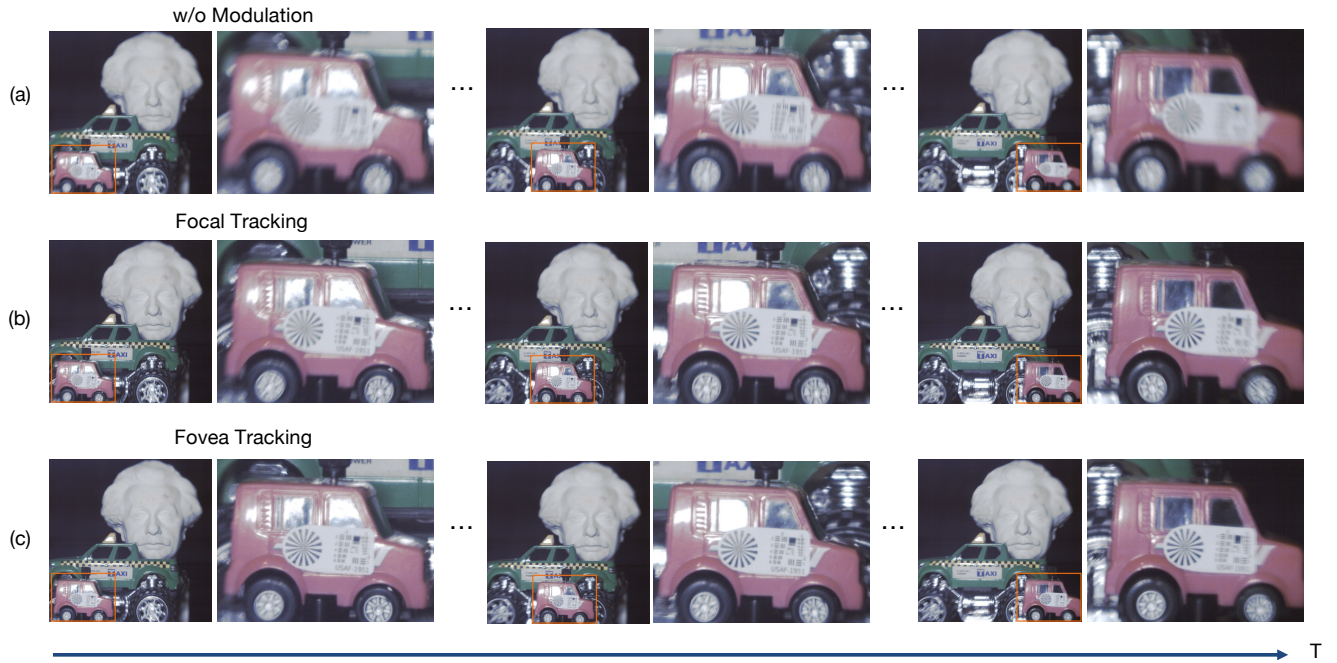


Fig. 19. Foveated Imaging for Object Tracking. A car moves at 1 mm/s on a translation stage across the image. (a) Without modulation, the object appear blurred on the image edges. (b) With adaptive focus tracking, field curvature is corrected for better local imaging quality. (c) Fovea Stacking keeps the object sharp throughout the video.

to focus adjustment (either by shifting the lens or with a liquid tunable lens), the DPP's dynamic free-shape deformation offers superior flexibility for enhancing image quality across diverse applications. Analysis demonstrates the robustness of the proposed sharpness-based fusion method, while the neural network-based approach fails to adapt to the blur kernel in Fovea Stacking. By integrating foveated imaging with object detection or eye tracking, we enable smooth pursuit of moving targets, dynamically adjusting the imaging focus to maintain the object within the fovea region—opening new possibilities for applications such as surveillance and foveated virtual reality displays.

These contributions are validated on a functional hardware prototype compact enough for use beyond controlled lab environments. As DPP devices evolve toward greater miniaturization, this method could allow their integration into mobile devices—similar to liquid tunable lenses—which is not feasible with other reflective wavefront modulators. We believe this work highlights how evolving dynamically tunable optical components can help simplify optical systems and redefine the boundary between optics and computing in future imaging devices.

The system's ability to correct the aberration is limited by the maximum achievable deformation of the hardware, which could be overcome by cascading multiple thin DPPs [Rajaeipour et al. 2020]. Promising future directions include real-time dynamic fovea sweeping and adaptive saccadic imaging guided by scene content. Since foveated imaging projects a 3D scene into multiple 2D views, finer 3D scene details can be reconstructed from foveated images

without moving the camera. For applications like extended depth of field imaging, optimal imaging parameters can be determined based on current 3D world estimates, enabling adaptive saccadic imaging.

REFERENCES

- Kaustubh Banerjee, Pouya Rajaeipour, Çağlar Ataman, and Hans Zappe. 2018. Optofluidic adaptive optics. *Applied optics* 57, 22 (2018), 6338–6344.
- Thomas Bifano. 2011. MEMS deformable mirrors. *Nature photonics* 5, 1 (2011), 21–23.
- Loz Blain. 2021. Xiaomi to launch the world's first liquid lens smartphone camera. <https://newatlas.com/mobile-technology/xiaomi-mi-mix-liquid-lens/>. Accessed: May 19, 2025.
- David S Bolme, J Ross Beveridge, Bruce A Draper, and Yui Man Lui. 2010. Visual object tracking using adaptive correlation filters. In *2010 IEEE computer society conference on computer vision and pattern recognition*. IEEE, 2544–2550.
- Jose Antonio Boluda, Fernando Pardo, Tomas Kayser, Juan J Pérez, and Joan Pelechano. 1996. A new foveated space-variant camera for robotic applications. In *Proceedings of Third International Conference on Electronics, Circuits, and Systems*, Vol. 2. IEEE, 680–683.
- Guillem Carles, Shouqian Chen, Nicholas Bustin, James Downing, Duncan McCall, Andrew Wood, and Andrew R Harvey. 2016. Multi-aperture foveated imaging. *Optics letters* 41, 8 (2016), 1869–1872.
- George Curatu, David V Wick, Don M Payne, Ty Martinez, Jamie Harriman, and James E Harvey. 2005. Wide field-of-view imaging system using a liquid crystal spatial light modulator. In *Current Developments in Lens Design and Optical Engineering VI*, Vol. 5874. SPIE, 80–86.
- Yuanshen Guan, Ruikang Xu, Mingde Yao, Lizhi Wang, and Zhiwei Xiong. 2023. Mutual-guided dynamic network for image fusion. In *Proceedings of the 31st ACM international conference on multimedia*. 1779–1788.
- Karen M Hampson, Raphaël Turcotte, Donald T Miller, Kazuhiro Kurokawa, Jared R Males, Na Ji, and Martin J Booth. 2021. Adaptive optics for high-resolution imaging. *Nature Reviews Methods Primers* 1, 1 (2021), 68.
- Jamie Harriman, Sebastian Gauza, Shin-Tson Wu, David Wick, Brett Bagwell, Ty Martinez, Don Payne, and Steven Serati. 2006. Transmissive spatial light modulators with high figure-of-merit liquid crystals for foveated imaging applications. In *Liquid Crystal Materials, Devices, and Applications XI*, Vol. 6135. SPIE, 80–92.

- Feng Huang, He Ren, Xianyu Wu, and Pengfei Wang. 2021. Flexible foveated imaging using a single Risley-prism imaging system. *Optics Express* 29, 24 (2021), 40072–40090.
- Wei Huang and Zhongliang Jing. 2007. Evaluation of focus measures in multi-focus image fusion. *Pattern recognition letters* 28, 4 (2007), 493–500.
- Sujit Kuthirummal, Hajime Nagahara, Changyin Zhou, and Shree K Nayar. 2010. Flexible depth of field photography. *IEEE transactions on pattern analysis and machine intelligence* 33, 1 (2010), 58–71.
- Chao Liu, Yi Zheng, Rong-Ying Yuan, Zhao Jiang, Jin-Bo Xu, You-Ran Zhao, Xin Wang, Xiao-Wei Li, Yan Xing, and Qiong-Hua Wang. 2023. Tunable liquid lenses: emerging technologies and future perspectives. *Laser & Photonics Reviews* 17, 11 (2023), 2300274.
- Luc. 2022. Third-order aberration theory. <https://www.thepulsar.be/article/-devoptical-part-14--third-order-aberration-theory> Accessed: May 19, 2025.
- Ty Martinez, David V Wick, and Sergio R Restaino. 2001. Foveated, wide field-of-view imaging system using a liquid crystal spatial light modulator. *Optics Express* 8, 10 (2001), 555–560.
- Daniel Miao, Oliver Cossairt, and Shree K Nayar. 2013. Focal sweep videography with deformable optics. In *IEEE International Conference on Computational Photography (ICCP)*. IEEE, 1–8.
- Mulun Na, Hector A Jimenez Romero, Xinge Yang, Jonathan Klein, Dominik L Michels, and Wolfgang Heidrich. 2024. End-to-end optimization of fluidic lenses. In *SIGGRAPH Asia 2024 Conference Papers*. 1–10.
- Shree K Nayar and Yasuo Nakagawa. 1994. Shape from focus. *IEEE Transactions on Pattern analysis and machine intelligence* 16, 8 (1994), 824–831.
- Yifan Peng, Qilin Sun, Xiong Dun, Gordon Wetzstein, Wolfgang Heidrich, and Felix Heide. 2019. Learned large field-of-view imaging with thin-plate optics. *ACM Trans. Graph.* 38, 6 (2019), 219–1.
- P Mirajkar Pradnya and D Ruikar Sachin. 2013. Wavelet based image fusion techniques. In *2013 international conference on intelligent systems and signal processing (ISSP)*. IEEE, 77–81.
- Pouya Rajaeipour, Kaustubh Banerjee, Alex Dorn, Hans Zappe, and Çağlar Ataman. 2020. Cascading optofluidic phase modulators for performance enhancement in refractive adaptive optics. *Advanced Photonics* 2, 6 (2020), 066005–066005.
- Pouya Rajaeipour, Martin Sauther, Kaustubh Banerjee, Hans Zappe, and Çağlar Ataman. 2021. Seventh-order wavefront modulation with a gravity-neutral optofluidic deformable phase plate. *Journal of Optical Microsystems* 1, 3 (2021), 034502–034502.
- Giulio Sandini and Giorgio Metta. 2003. Retina-like sensors: motivations, technology and applications. In *Sensors and sensing in biology and engineering*. Springer, 251–262.
- Zheng Shi, Xiong Dun, Haoyu Wei, Siyu Dong, Zhanshan Wang, Xinbin Cheng, Felix Heide, and Yifan Peng. 2024. Learned multi-aperture color-coded optics for snapshot hyperspectral imaging. *ACM Transactions on Graphics (TOG)* 43, 6 (2024), 1–11.
- Haozhe Si, Bin Zhao, Dong Wang, Yunpeng Gao, Mulin Chen, Zhigang Wang, and Xuelong Li. 2023. Fully self-supervised depth estimation from defocus clue. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 9140–9149.
- Vincent Sitzmann, Steven Diamond, Yifan Peng, Xiong Dun, Stephen Boyd, Wolfgang Heidrich, Felix Heide, and Gordon Wetzstein. 2018. End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging. *ACM Transactions on Graphics (TOG)* 37, 4 (2018), 1–13.
- Jianguo Sun, Qilong Han, Liang Kou, Liguang Zhang, Kejia Zhang, and Zilong Jin. 2018. Multi-focus image fusion algorithm based on Laplacian pyramids. *Journal of the Optical Society of America A* 35, 3 (2018), 480–490.
- Qilin Sun, Congli Wang, Fu Qiang, Dun Xiong, and Heidrich Wolfgang. 2021. End-to-end complex lens design with differentiable ray tracing. *ACM Trans. Graph.* 40, 4 (2021), 1–13.
- Qilin Sun, Jian Zhang, Xiong Dun, Bernard Ghanem, Yifan Peng, and Wolfgang Heidrich. 2020. End-to-end learned, optically coded super-resolution SPAD camera. *ACM Transactions on Graphics (TOG)* 39, 2 (2020), 1–14.
- Larry N Thibos, Raymond A Applegate, James T Schwiegerling, and Robert Webb. 2002. Standards for reporting the optical aberrations of eyes. , S652–S660 pages.
- Simon Thiele, Kathrin Arzenbacher, Timo Gissibl, Harald Giessen, and Alois M Herkommer. 2017. 3D-printed eagle eye: Compound microlens system for foveated imaging. *Science advances* 3, 2 (2017), e1602655.
- Ethan Tseng, Ali Mosleh, Fahim Mannan, Karl St-Arnaud, Avinash Sharma, Yifan Peng, Alexander Braun, Derek Nowrouzezahrai, Jean-Francois Lalonde, and Felix Heide. 2021. Differentiable compound optics and processing pipeline optimization for end-to-end camera design. *ACM Transactions on Graphics (TOG)* 40, 2 (2021), 1–19.
- Congli Wang, Ni Chen, and Wolfgang Heidrich. 2022. do: A differentiable engine for deep lens design of computational imaging systems. *IEEE Transactions on Computational Imaging* 8 (2022), 905–916.
- Congli Wang, Qiang Fu, Xiong Dun, and Wolfgang Heidrich. 2018. Megapixel adaptive optics: towards correcting large-scale distortions in computational cameras. *ACM Transactions on Graphics (TOG)* 37, 4 (2018), 1–12.
- Congli Wang, Qiang Fu, Xiong Dun, and Wolfgang Heidrich. 2019. Quantitative phase and intensity microscopy using snapshot white light wavefront sensing. *Scientific Reports* 9, 1 (2019), 13795.
- Ning-Hsu Wang, Ren Wang, Yu-Lun Liu, Yu-Hao Huang, Yu-Lin Chang, Chia-Ping Chen, and Kevin Jou. 2021. Bridging unsupervised and supervised depth from focus via all-in-focus supervision. In *Proceedings of the IEEE/CVF international conference on computer vision*. 12621–12631.
- Wencheng Wang and Faliang Chang. 2011. A Multi-focus Image Fusion Method Based on Laplacian Pyramid. *J. Comput.* 6, 12 (2011), 2559–2566.
- David V Wick, Ty Martinez, Sergio R Restaino, and BR Stone. 2002. Foveated imaging demonstration. *Optics Express* 10, 1 (2002), 60–65.
- Xinge Yang, Qiang Fu, and Wolfgang Heidrich. 2024a. Curriculum learning for ab initio deep learned refractive optics. *Nature communications* 15, 1 (2024), 6572.
- Xinge Yang, Matheus Souza, Kunyi Wang, Praneeth Chakravarthula, Qiang Fu, and Wolfgang Heidrich. 2024b. End-to-end hybrid refractive-diffractive lens design with differentiable ray-wave model. In *SIGGRAPH Asia 2024 Conference Papers*. 1–11.
- Yiqian Yang, Andrew Forbes, and Liangcai Cao. 2023. A review of liquid crystal spatial light modulators: devices and applications. *Opto-Electronic Science* 2, 8 (2023), 230026–1.
- Nanfeng Yu, Patrice Genevet, Mikhail A Kats, Francesco Aieta, Jean-Philippe Tetienne, Federico Capasso, and Zeno Gaburro. 2011. Light propagation with phase discontinuities: generalized laws of reflection and refraction. *science* 334, 6054 (2011), 333–337.
- Rabia Zafar, Muhammad Shahid Farid, and Muhammad Hassan Khan. 2020. Multi-focus image fusion: algorithms, evaluation, and a library. *Journal of Imaging* 6, 7 (2020), 60.
- Xingchen Zhang. 2021. Deep learning-based multi-focus image fusion: A survey and a comparative study. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44, 9 (2021), 4819–4838.
- Yu Zhang, Yu Liu, Peng Sun, Han Yan, Xiaolin Zhao, and Li Zhang. 2020. IFCNN: A general image fusion framework based on convolutional neural network. *Information Fusion* 54 (2020), 99–118.
- Xiaoxia Zhao, Yongjun Xie, and Wei Zhao. 2008a. Broadband and wide field of view foveated imaging system in space. *Optical Engineering* 47, 10 (2008), 103202–103202.
- Xiaoxia Zhao, Yongjun Xie, and Wei Zhao. 2008b. Wide field-of-view foveated imaging system. *Chinese Optics Letters* 6, 8 (2008), 561–563.