This ICCV paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

A Differentiable Wave Optics Model for End-to-End Computational Imaging System Optimization

Chi-Jui Ho Yash Belhe Steve Rotenberg Ravi Ramamoorthi Tzu-Mao Li Nicholas Antipa
University of California, San Diego
{chh009, ybelhe, srotenberg, ravir, tzli, nantipa}@ucsd.edu

Abstract

End-to-end optimization, which integrates differentiable optics simulators with computational algorithms, enables the joint design of hardware and software for data-driven imaging systems. However, existing methods usually compromise physical accuracy by neglecting wave optics or off-axis effects due to the high computational cost of modeling both aberration and diffraction. This limitation raises concerns about the robustness of optimized designs. In this paper, we propose a differentiable optics simulator that accurately and efficiently models aberration and diffraction in compound optics and allows us to analyze the role and impact of diffraction in end-to-end optimization. Experimental results demonstrate that compared with ray-optics-based optimization, diffraction-aware optimization improves system robustness to diffraction blur. Through accurate wave optics modeling, we also apply the simulator to optimize the Fizeau interferometer and freeform optics elements. These findings underscore the importance of accurate wave optics modeling in robust end-to-end optimization. Our code is publicly available at: https://github.com/JerryHoTaiwan/DeepWaveOptics

1. Introduction

The interdependence between optics and downstream algorithms is pivotal in imaging system design. To leverage this interdependence and achieve joint designs, end-to-end differentiable models, which incorporate a differentiable optics simulator and a computer vision algorithm, have been applied to simultaneously optimize hardware and software across a range of vision tasks [5, 18, 22–25, 31, 32]. Given an image, the differentiable simulator models the corresponding measurement taken by the optics system, and the computer vision algorithm extracts semantic information. With a differentiable simulator and algorithm, a loss function scores task performance and drives the optimization of the optics and algorithm parameters via backpropagation.

A notable challenge in end-to-end optimization is incorporating wave optics effects in large field-of-view (FoV) and analyzing how the fidelity of optics simulation impacts overall system optimization. Realistic modeling requires accounting for diffraction across the entire sensor, which is computationally expensive. Thus, many designs neglect diffraction and adopt ray optics [5, 24, 32]. Some simulators consider simplified diffraction using thin-phase surfaces [22, 29] or shift-invariance [2, 9, 23], limiting applicability to multi-element or compound optics. Although recent frameworks model more realistic wave optics [4, 33], their accuracy and efficiency in different configurations remain questionable, and the significance of wave optic effects on system optimization remains an open problem.

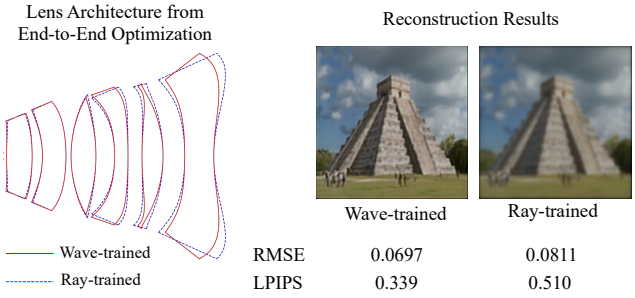


Figure 1. **End-to-end optimized lens architectures and reconstruction models using ray and wave optics.** By taking diffraction into account, our wave-trained model yields sharper reconstruction results than the baseline using ray optics.

In this paper, we propose an accurate, efficient, and differentiable optics simulator, which uses ray tracing with the Rayleigh-Sommerfield integral [7] to model diffraction and off-axis aberrations in compound optical systems without thin-phase or paraxial approximations. To effectively model diffraction in large FoVs, we use an interpolation method to approximate the measurements with a subset of point spread functions (PSFs). By providing accurate and efficient wave optics rendering, the proposed simulator enables us to incorporate diffraction into end-to-end optimization and analyze its role and impact on imaging system design.

Unlike systems optimized solely under ray optics assumptions, our wave optics model guides the system to a

design with reduced diffraction effects. An example of lens architecture and system performance optimized by ray and wave optics is shown in Fig. 1.

Our contributions are

- We propose a differentiable model that accurately accounts for aberration and diffraction in compound optical systems. With efficient rendering, the model is compatible with end-to-end optimization.
- We analyze the role of diffraction in end-to-end design. Neglecting diffraction leads to suboptimal lens and algorithm configurations. Conversely, by accurately modeling diffraction, our model attains superior solutions.
- The proposed simulator is applicable to a wide range of wave-optics-based imaging systems, including interferometric setups and freeform optical systems.

2. Related Work

2.1. End-to-End Optimization

Conventional lens design optimizes a merit function that combines lens properties with transfer function quality [16]. However, such merit functions do not necessarily correlate with computer vision task performance [6, 32]. End-to-end optimization addresses this by directly optimizing task performance, jointly refining hardware and software. Leveraging differentiable optics simulators and inference algorithms on large datasets [5, 9, 24, 32], this approach provides data-driven designs that capture interdependencies among optics, algorithms, and tasks [6].

This paradigm has been applied to image reconstruction [6, 13, 14, 18, 22, 23] and restoration [8, 35]. Sitzmann *et al.* extend the depth of field in computational cameras [23], while Peng *et al.* achieve high-FoV image reconstruction [18]. Shi *et al.* combine diffractive optics with point-PSF-aware neural networks to recover occluded scenes [22]. Beyond reconstruction, this strategy has also been applied to semantic tasks: Baek *et al.* jointly optimize diffractive elements and networks for hyperspectral depth sensing [2]; Kellman *et al.* design coded-illumination patterns and unrolled networks for phase recovery [11]; Pidhorskyi *et al.* develop a differentiable ray tracer for depth-of-field-aware intensity recovery [19]; Yang *et al.* optimize off-axis aberrations for image classification [32]; and Cote *et al.* co-optimize lens materials and structures for object detection [5]. These works demonstrate how end-to-end optimization yields task-specific optics–algorithm co-design.

2.2. Balancing Accuracy and Efficiency in Differentiable Optics Simulation

Zemax is an industry-standard tool for accurate wave-optics modeling using Huygens' principle, but its slow computational speed restricts its use in end-to-end optimization [16], which requires gradient propagation to model complex optical–semantic relationships [27]. To mitigate the cost in large FoV differentiable rendering, simplified physics models are usually adopted, such as thin-lens modeling [18, 22] and geometric ray tracing [5, 24, 27], to compute PSFs in end-to-end optimization. However, the former is limited to a single thin lens, and the latter neglects wave effects.

It is also common to model simplified wave-optical effects. Sitzmann *et al.* use Fresnel propagation for shift-invariant diffractive optics [23]; He *et al.* compute PSFs with diffraction theory in shift-variant systems [9]; and Tseng *et al.* replace the full pipeline with a neural PSF renderer [26]. All these assumptions limit their applicability to compound optical systems. Chen *et al.* [4] simulate diffraction with ray tracing but neglect magnitude variations with propagation distance, which become inaccurate when modeling defocus. Wei *et al.* [29] and Yang *et al.* [33] use the angular spectrum method (ASM) for free-space propagation, which requires high sampling density when modeling defocus. A similar limitation arises in field tracing [30], which unifies different propagation models but still relies on ASM, while the generalized Debye integral [28] accelerates focal-field computation via a homeomorphic Fourier transform yet remains constrained to low Fresnel number regimes. These considerations motivate our differentiable ray–wave framework, which explicitly models diffraction from rays and examines the impact of accurate wave modeling on end-to-end optimization.

3. Differentiable Optics Model

Our differentiable simulator is designed to accurately and efficiently capture both aberration and diffraction in compound optics systems, making it a robust rendering model and providing scalable end-to-end optimization.

An overview of our differentiable hybrid ray-wave imaging simulator is shown in Fig. 2. Given a point light source at $\mathbf{x} = (x, y, z)$ and an optical system with sequential refractive surfaces, our model incorporates a differentiable ray tracer [27] and Rayleigh-Sommerfield integral [7] to account for wave optics effects in PSF $h(\mathbf{u}|\mathbf{x})$, where $\mathbf{u}$ denotes sensor pixel position. We describe PSF rendering in detail in Sec. 3.1. Furthermore, given scene intensity $b(\mathbf{x})$, the resulting measurement $I(\mathbf{u})$ is derived from the superposition integral of incoherent PSFs [4]:

$$I(\mathbf{u}) = \int b(\mathbf{x}) h(\mathbf{u}|\mathbf{x}) d\mathbf{x}. \tag{1}$$

Directly computing Eq. (1) across the full FoV is computationally intensive, requiring full-resolution PSF rendering for every point source. To address this, we develop an efficient interpolation method that balances accuracy and computational cost. The approach involves sampling a subset of PSFs and using interpolation to approximate the full measurement by convolving the subset PSFs with their corre-

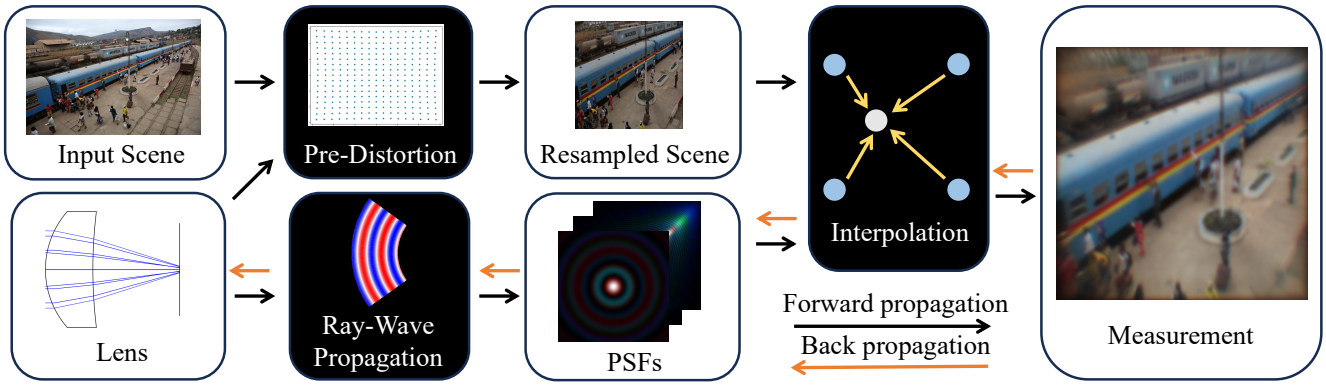


Figure 2. **Our proposed differentiable wave optics simulator.** Given an input scene and lens configuration, we first resample the scene based on the lens' pre-distortion map. Next, we generate diffraction-aware PSFs using our wave optics simulator. Finally, we interpolate the convolution of the resampled scene with the PSFs to obtain our final measurement. During lens optimization, measurement gradients are back-propagated to the lens parameters.

sponding sub-scene intensities [3] (details in Sec. 3.2). By efficiently modeling diffraction, off-axis aberrations, and geometric distortions, this approach enhances the robustness of data-driven lens design and enables deeper analysis of wave optics effects in imaging systems.

3.1. PSF Rendering

A conceptual flow of our wave optics model is illustrated in Fig. 3. To compute a PSF, we first use geometric ray tracing to sample the wavefront map in the exit pupil, and then propagate the complex field of the wavefront map to the sensor plane. In ray tracing, we use Newton's Method [24, 27] to calculate the intersections between rays and surfaces and use Snell's Law to model refractions. The wavefront map is then calculated in the reference sphere, whose center and radius are determined by the intersection between the principal ray (the ray that passes through the pupil center) and the sensor plane, and the distance between the exit pupil (XP) and the sensor, respectively. Specifically, we approximate the XP using paraxial rays across all incident angles, consistent with prior work [13, 16, 33].

The task thus reduces to computing the amplitude and phase of the complex field on the reference sphere. Since the exit pupil is an image of the aperture stop, we model the amplitude by the square root of the aperture stop transmittance. The phase at the reference sphere is determined by the optical path length (OPL) δ calculated by

$$\delta = \int_C n(s)ds, \tag{2}$$

where $n(s)$ is the 3D refractive index of the system and C is the path that a given ray takes from the light source to the reference sphere [10].

It is notable that the phase of complex values across the reference sphere, called the *wavefront error map*, reflects the degree of focusing [21]. When the system is in-

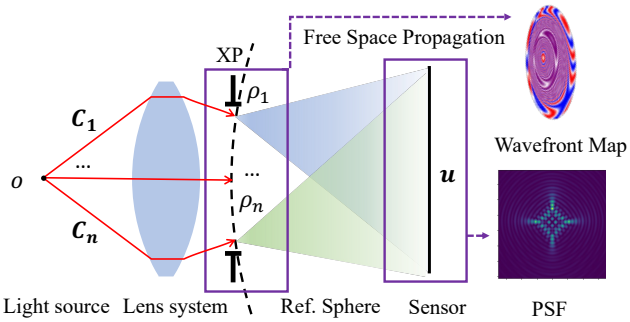


Figure 3. **Our wave optics simulator.** We trace rays emitted from a point source o to the reference sphere on the system's exit pupil, and compute intersections $\{\boldsymbol{\rho}_i\}$ and associated phase on a wavefront map. We then perform free-space propagation toward the sensor to generate a PSF. XP: Exit Pupil. Ref: Reference.

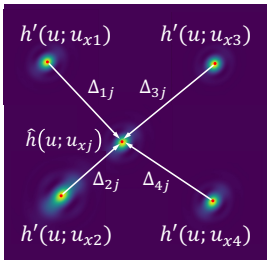


Figure 4. **Approximating unsampled PSFs.** Our system first samples PSFs $h'(u;\cdot)$ on a regular grid. Next, by exploiting the isoplanatic property, it approximates off-grid PSFs $\hat{h}(u;\cdot)$ by interpolating shifted and scaled versions of nearest samples $h'(u;\cdot)$.

focus, the reference sphere exactly matches the wavefront, and the phase is constant on the sphere. Otherwise, the mismatch between the actual wavefront and the reference sphere causes phase variations across the reference sphere. Moreover, compared with the planar pupil field used by ASM-based modeling [29, 31], the spherical structure effectively reduces the phase variation and hence alleviates

the sampling requirement. In other words, we choose the reference sphere to model the wavefront error map because of its interpretability, efficiency, and compatibility with our propagation model, but the choice of the reference geometry is arbitrary and depends on the propagation model [16, 21].

Consequently, for a ray piercing the reference sphere at $\boldsymbol{\rho_i} = (\rho_{x_i}, \rho_{y_i}, \rho_{z_i})$, we model the complex field by

$$v(\boldsymbol{\rho}_i) = a_i \exp(jk\delta_i), \tag{3}$$

where a_i is the amplitude, k is the wave number, $j = \sqrt{-1}$, and δ_i is the optical path length.

As shown in Fig. 3, the propagation from the reference sphere to the sensor is in free space. The total intensity, $h(\mathbf{u})$, at sensor coordinate $\mathbf{u}$ is computed by the Rayleigh-Sommerfeld integral [7], which we Monte-Carlo evaluate with N coherent rays by

$$h(\mathbf{u}) = \frac{1}{N\lambda^2}\left|\sum_{i=1}^{N} v(\boldsymbol{\rho}_i)\frac{\exp(jk|\vec{r}_{u,i}|)}{|\vec{r}_{u,i}|}\cos(\theta_{u,i})\right|^2, \tag{4}$$

where $\vec{r}_{u,i}$ denotes the vector from $\boldsymbol{\rho}_i$ to sensor coordinate $\mathbf{u}$, and $\theta_{u,i}$ is the angle between $\vec{r}_{u,i}$ and the normal vector of the reference geometry at $\boldsymbol{\rho}_i$. To accelerate computation, we vectorize the operations; however, because full vectorization can exceed memory limits, we apply checkpointing in PyTorch [17] to alleviate this issue.

3.2. Approximating Superposition Integral

Although we can render PSFs with wave optics effects, the high computational costs make it challenging to exhaustively compute all PSFs. A common way to alleviate this cost is to assume the system is shift-invariant and approximate Eq. (1) with a single convolution between the on-axis PSF and scene intensities [7]. However, this assumption is overly restrictive as it does not model common off-axis aberrations such as coma, astigmatism, and field curvatures.

Therefore, we assume that PSFs are locally *isoplanatic*; the system is shift-invariant over a sufficiently small area. This allows us to sample a small subset of PSFs and approximate the superposition integral through a sequence of convolutions, thereby saving computational costs while maintaining the ability to model off-axis aberrations.

To facilitate the derivation, we parameterize scene intensities $b(\mathbf{x})$ and PSFs $h(\mathbf{u};\mathbf{x})$ in terms of sensor coordinates $\{\mathbf{u}\}$ as follows. Given a world coordinate $\mathbf{x}$ and lens distortion function $d(\cdot)$, we compute the intersection $\mathbf{u_x} = d(\mathbf{x})$ between the non-paraxial principal ray emitting from $\mathbf{x}$ and the sensor plane. Because the function is one-to-one, $b(\mathbf{x})$ and $h(\mathbf{u};\mathbf{x})$ can be re-parameterized as $b'(\mathbf{u_x})$ and $h'(\mathbf{u};\mathbf{u_x})$, respectively. An example of distorted coordinates is visualized in Fig. 2. Because the distortion function $d(\cdot)$ only determines the input scene content, we only consider it in the inference, but not back-propagation.

Fig. 4 shows an example of approximating a PSF originating from an *unsampled* world coordinate $\mathbf{x_j}$ according to PSFs $\{h(\mathbf{u};\mathbf{u_{xi}})\}$ originating from sampled world coordinates $\{\mathbf{x}_i\}$. For an unsampled PSF centered at $\mathbf{u_{xj}}$, we model it as the weighted sum of the known neighboring PSFs, which are aligned to the same location:

$$\widehat{h}(\mathbf{u};\mathbf{u_{xj}}) = \sum_i w_i(\mathbf{u_{xj}})h'(\mathbf{u}-\Delta_{ij};\mathbf{u_{xi}}), \tag{5}$$

where $\Delta_{ij} = \mathbf{u_{xj}} - \mathbf{u_{xi}}$ is the center-to-center distance, in the sensor space, between the sampled PSF i and unsampled PSF j. $w_i(\mathbf{u_{xj}})$ determines the weight of the sampled PSF i when approximating the unsampled PSF centered at $\mathbf{u_{xj}}$.

Therefore, we rewrite Eq. (1) by substituting the general form for the shift-varying PSFs found in Eq. (5):

$$\begin{aligned} I(\mathbf{u}) &= \sum_{\mathbf{u_x}} b'(\mathbf{u_x}) \sum_i w_i(\mathbf{u_x})h'(\mathbf{u}+\mathbf{u_{xi}}-\mathbf{u_x};\mathbf{u_{xi}}) \\ &= \sum_i \sum_{\mathbf{u_x}} b_i''(\mathbf{u_x})h'(\mathbf{u}+\mathbf{u_{xi}}-\mathbf{u_x};\mathbf{u_{xi}}), \end{aligned} \tag{6}$$

where $b_i''(\mathbf{u_x}) = b'(\mathbf{u_x})w_i(\mathbf{u_x})$ represents the weighted latent image, which consists of input scene intensities distorted by the lens distortion curve and weighted by $w_i(\cdot)$.

We observe that Eq. (6) is a sum of convolutions between the shifted version of sampled PSFs and the corresponding weighted latent image:

$$\begin{aligned} I(\mathbf{u}) &= \sum_i \sum_{\mathbf{u_x}} b_i''(\mathbf{u_x})h_i(\mathbf{u}-\mathbf{u_x}) \\ &= \sum_i b_i'' * h_i \end{aligned} \tag{7}$$

where $h_i(\mathbf{u}) = h'(\mathbf{u}+\mathbf{u_i};\mathbf{u_{xi}})$. Fig. 5 illustrates an example of how we pair weighted images and PSFs, convolve them with each other, and sum up the convolved images to compute the measurement. Notably, because $v(\boldsymbol{\rho}_i)$ is obtained by differentiable ray tracing [27], and the operations from (4) to (7) are all differentiable, the entire pipeline remains differentiable. This property enables precise modeling of how lens configurations interact with wave-optics effects to produce the measurements.

Although Zemax also uses Huygens' Principle to model wave propagation and serves as an industrial level baseline [16], it requires on-grid sampling to model the wavefront map, which limits the efficiency. The differentiability of Zemax is also limited to the given merit functions, while our simulator can be integrated with arbitrary differentiable algorithms. In the subsequent section, we incorporate this differentiable wave optics simulator into computer vision algorithms, allowing analysis of the impact of wave optics effects on optical systems tailored for vision tasks.

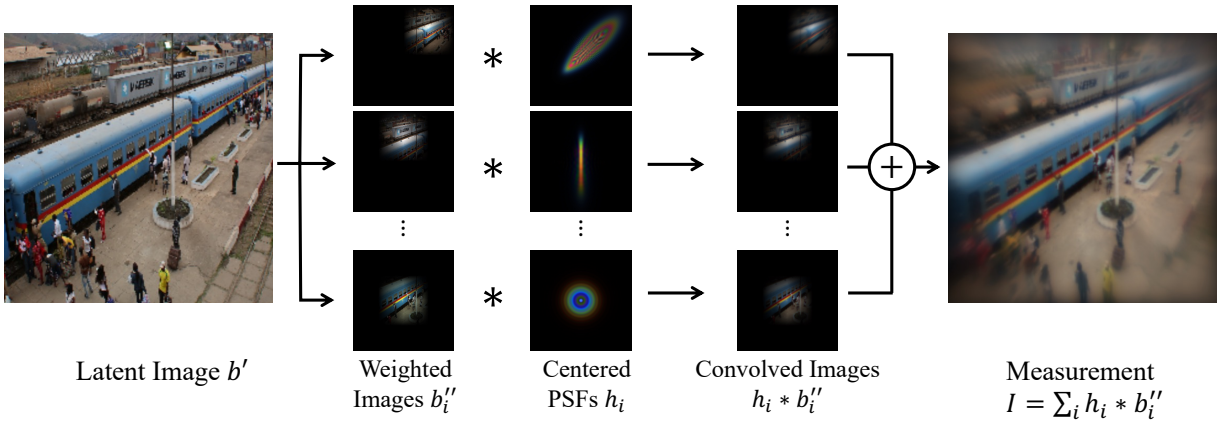


Figure 5. **Rendering measurement with a subset of PSFs.** Given a latent image b', we first generate weighted images b_i''. Next, we generate PSFs h_i at the centers of weighted images and pair them with corresponding PSFs. Finally, we convolve weighted images and PSFs ($h_i * bi''$) and sum them up to obtain the measurement I.

4. Experiments

With the simulator, we conduct joint optimization of optics systems and scene reconstruction algorithms, with a focus on analyzing the role of diffraction in end-to-end optimization. To the best of our knowledge, it is an unexplored experimental flow to analyze the requirements of physical accuracy in end-to-end optimization. We also analyze the rendering and interpolation accuracy of our simulator and extend the simulator to interferometry and freeform optics.

4.1. PSF Rendering

In Fig. 6, we present monochromatic PSFs (wavelength: 532 nm) generated by our simulator alongside those from existing methods [4, 16, 29, 33] under various conditions: On-axis PSFs for an in-focus and out-of-focus Cooke Triplet lens, and off-axis PSFs at 35° and 40° from a singlet lens. Because Zemax computes Huygens PSFs using the Rayleigh–Sommerfeld integral [7, 16], the most general scalar diffraction model, we use Zemax-Huygens results as the reference. For each method, we report similarity to this reference using the structural similarity index (SSIM) and evaluate efficiency with ray count and computational time.

Overall, our method achieves superior accuracy and efficiency. While our simulator requires more rays than Zemax-FFT in the in-focus case, it attains higher accuracy with shorter runtime, underscoring both precision and efficiency. Notably, ASM-based methods [29, 33] are sensitive to defocus: as defocus increases, phase variations across the pupil plane and propagation kernel become extremely rapid. Since ASM discretizes these on a 2D grid, it struggles to capture such variations, reducing accuracy and efficiency. In contrast, our renderer supports off-grid wavefront maps that directly represent ray distributions, enabling efficient modeling of wave propagation. Although Chen et al. [4] also allows flexible ray distributions, their wave model does not account for magnitude changes from $|\vec{r}_{u,i}|$ in Eq. 4, instead projecting $|\vec{r}_{u,i}|$ onto ray directions. This approximation fails to capture magnitude variations across large spot sizes under defocus. Furthermore, defocus makes the Airy disk, commonly used to evaluate diffraction in ideal lenses, unreliable for modeling wave effects. Our results demonstrate that the proposed simulator is more robust, accurate, and efficient for defocused and large-FoV systems, conditions frequently encountered in end-to-end optimization. Importantly, our method is not merely reproducing the Huygens PSFs from Zemax but is more efficient and directly compatible with differentiable algorithms.

4.2. System Optimization Setup

In our imaging rendering process, we simulate beam propagation across the red, green, and blue light channels, compute the corresponding measurements for each wavelength, and then apply the Bayer filter to subsample these measurements. This results in blurred and mosaicked data.

We perform both ray-based and wave-based end-to-end optimization to jointly design lens systems and a U-Net [20] for scene reconstruction from system measurements. To compare their robustness to diffraction effects, we use wave optics in the evaluation. Input scenes are drawn from the DIV2K dataset [1], and lens configurations include variations in aperture radii and complexity, encompassing singlet, triplet, and six aspheric lenses.

For optimization, we utilize the Adam optimizer [12] to adjust both the network and lens parameters. The loss function includes root-mean-square error (RMSE) and perceptual loss (LPIPS) [34] between the normalized input scene intensities and the reconstructed results. To keep a consistent FoV for fair comparisons, whenever the focal length varies, we adjust the sensor size accordingly.

In addition to assessing reconstruction with RMSE and LPIPS, we use two metrics to quantify the disparity between ray- and wave-trained lenses: The mismatch between their F-numbers (MF) and the relative root mean squared error

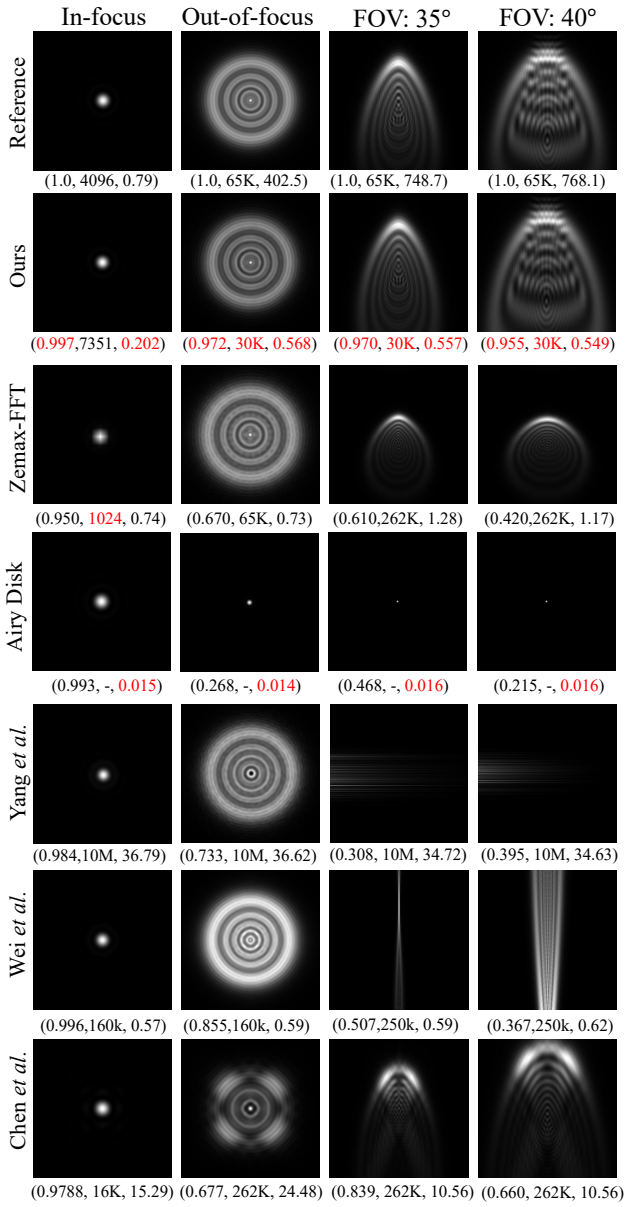


Figure 6. **PSFs rendered by different simulators under different conditions.** Unlike existing simulators [4, 16, 29, 33], ours avoids wavefront discretization and remains robust to defocus and large FoVs, achieving the highest accuracy and efficiency. The tuple (SSIM, ray count, time in sec.) highlights the best performance in red. As the Airy disk does not use ray-tracing, we skip its ray count and do not compare its time with others. Zoom in for details.

(RRMSE) of optimizable variables. All experiments were implemented on an Nvidia A40 GPU using PyTorch [17].

4.3. Demosaicking and Reconstruction

We summarize the reconstruction results in Table 1, which are consistently evaluated using wave optics. Notably, with lenses having a 0.1 mm aperture radius, wave-training and ray-training yield different configurations and reconstruction performance. In Fig. 7, we visualize ray- and wave-trained lens configurations and associated PSFs and reconstructions at different testing situations. As shown in Fig. 7 (a) and (b), the wave-trained lens changes its architecture to shorten the focal length and weaken diffraction, while the ray-trained lens focuses on minimizing RMS spot size.

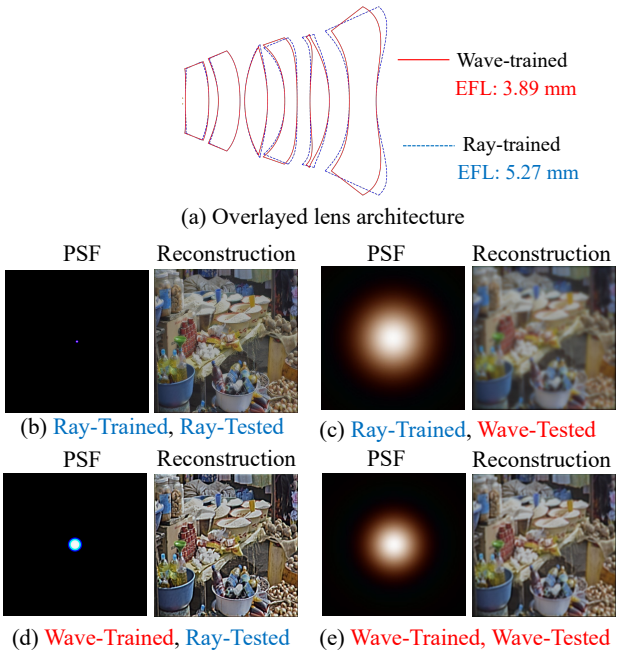


Figure 7. **Ray- vs. wave-trained systems.** (a) Lens architecture of wave- and ray-trained systems. The ray-trained system minimizes geometric spot size (b) but neglects diffraction blur (c). The wave-trained system has a larger geometric spot size (d), but a lower effective focal length (EFL) to control diffraction, yielding better reconstruction (e). PSF size: 0.044 mm^2.

Although the ray-trained system achieves a smaller geometric spot size, as shown in Fig. 7 (b) and (d), it fails to account for diffraction blur. When evaluated by accurate wave modeling, as shown in Fig. 7 (c) and (e), both PSF quality and reconstruction performance degrade. In contrast, while the wave-trained system slightly sacrifices geometric spot size, its optimized lens architecture effectively mitigates diffraction, the actual PSF-limiting factor, enhancing diffraction-limited resolution and producing sharper reconstructions. This highlights the critical role of diffraction in end-to-end optimization and the risks of neglecting it.

Table 1 also shows that increasing the aperture radius from 0.1 to 0.3 mm reduces the mismatch between lens designs and the performance gap arising from different physics models. At a 0.1 mm aperture, the diffraction spot size significantly exceeds the geometric spot size, allowing the system to adjust its structure to balance aberration and diffraction effects. However, as the aperture increases, the system becomes aberration-limited, reducing the incentive to trade aberration performance for diffraction control.

Table 1. Reconstruction performance on wave-optics-rendered measurements (RMSE / LPIPS) and lens disparity.

AR	Training physics		MF	RRMSE
	Wave	Ray		
Singlet Lens				
0.1	**0.075 / 0.181**	0.089 / 0.451	1.11	5.1×10^{-3}
0.3	**0.065** / 0.076	0.063 / **0.073**	0.108	6.8×10^{-4}
Cooke Triplet Lens				
0.1	**0.106 / 0.265**	0.148 / 0.772	8.689	0.580
0.3	**0.104 / 0.230**	0.112 / 0.483	0.073	4.8×10^{-3}
Six Aspherical Lenses				
0.1	**0.085 / 0.368**	0.104 / 0.604	6.873	0.263
0.3	**0.067 / 0.173**	0.071 / 0.242	0.432	0.060

AR: Aperture radius (unit: mm)

Table 2. The SSIM between sparsely interpolated and reference measurement and time elapsed in interpolation.

Lens	FoV	Number of PSFs in interpolation			
		9	25	81	289
Singlet	5°	0.987	0.990	0.995	0.999
	15°	0.894	0.954	0.974	0.996
	30°	0.815	0.842	0.871	0.981
	Time	7.66	12.38	36.10	96.50
Cooke Triplet	5°	0.995	0.995	0.996	0.999
	15°	0.994	0.995	0.996	0.999
	30°	0.889	0.942	0.957	0.993
	Time	7.21	9.71	25.47	62.33
Six Aspheric	5°	0.998	0.998	0.998	0.999
	15°	0.996	0.996	0.997	0.997
	30°	0.998	0.998	0.998	0.999
	Time	8.70	13.68	37.03	104.06

Moreover, compared with the singlet lens, the Cooke triplet and six-asphere designs have higher structural flexibility and hence exhibit more variation in lens configurations.

We further investigate the impact of diffraction in the optimization of aberration-limited optics in Fig. 8. The experiments are conducted in a single lens at a 30° off-axis field point with a wavelength of 440 nm. As observed, despite structural differences between wave- (h_w) and ray-PSF (h_r), their spectra remain similar at low frequencies, where the energy of the natural image (I_N) spectrum is concentrated. Thus, their convolved sub-scenes, $h_w * I_N$ and $h_r * I_N$, exhibit negligible MSE. The MSE is only noticeable between measurements from inputs with rich high-frequency contents, such as $h_w * I_S$ and $h_r * I_S$, which are rare in existing datasets. As a result, with natural imaging datasets and aberration-limited systems, diffraction plays a minor role in end-to-end optimization.

4.4. Interpolation

To evaluate interpolation accuracy across different FoVs and lens complexities, we use measurements rendered with 969 PSFs, the maximum feasible under hardware limits, as the reference and compare them to interpolated measurements using 9–289 PSFs. As shown in Table 2, systems with larger FoV require more PSFs to reduce disparity due to stronger aberrations and reduced isoplanaticity. Conversely, systems with more complex lenses exhibit weaker aberrations and thus need fewer PSFs for accurate rendering. Table 2 also lists the computation time for interpolating a single image, showing that denser interpolation significantly increases cost. Thus, selecting the appropriate number of PSFs is critical to balance accuracy and efficiency, depending on FoV and lens complexity.

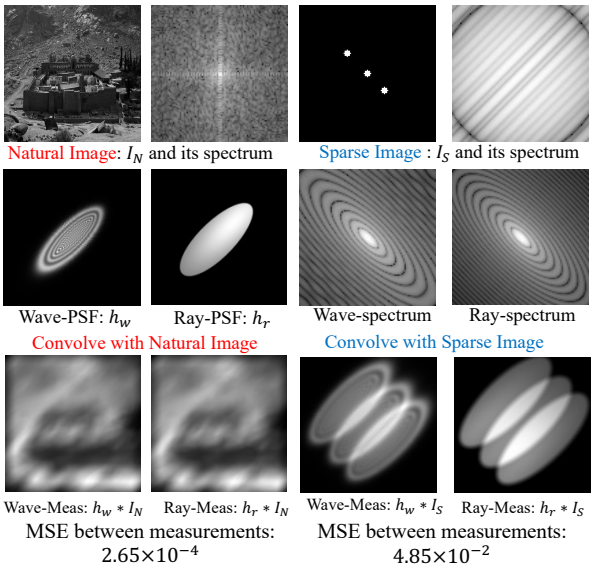
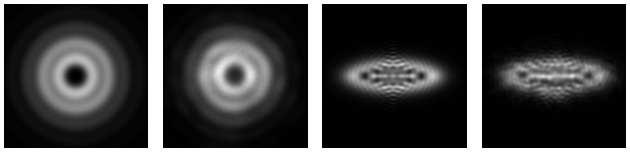


Figure 8. **Comparing ray and wave measurements in an aberration limited system.** The key spectral difference between ray- and wave-PSFs lies in high frequencies, affecting measurements only when the image has rich high-frequency components. Thus, for natural images, both systems receive similar training data and yield similar configurations. The MSE is measured using normalized measurement intensity.



(a) On-Sim. (b) On-Real. (c) Off-Sim. (d) Off-Real.

Figure 9. **Comparing simulated and real PSFs.** By sending monochromatic parallel beams into a physical lens, we measure real on- and off-axis (15°) PSFs (Real) and compare them with our simulated measurements (Sim.). Our simulator closely matches the real measurements by accurately modeling diffraction and aberration. PSF size: 0.217 (on) and 0.62 (off) mm^2.

4.5. Hardware Validation

We validate the physical accuracy of our simulator against real-world hardware implementations. In Fig. 9, we

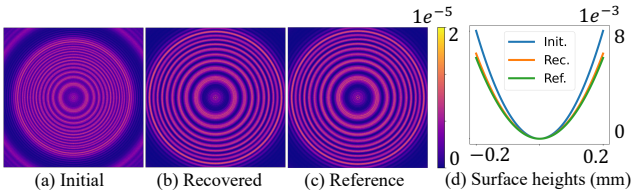


Figure 10. **Recovering a quadratic surface based on Fizeau interferometer measurements.** Setup: A coherent wavefront is reflected by a quadratic surface, and the resulting interference pattern is detected by the sensor. The interference pattern is determined by the surface geometry. By accurately modeling interference, our differentiable wave optics model results in accurate surface recovery (d). Sensor size: 1.6 mm^2.

send on-axis (0°) and off-axis (15°) parallel monochromatic beams (wavelength: 532 nm) through a plano-convex lens (model 011-1580) onto a sensor (UI-3882LE0M) to generate PSFs, which we then compare with simulated ones. As observed, our simulator accurately models the diffraction patterns and off-axis aberration, yielding similar structures in real and simulated PSFs. The SSIM between real and simulated PSFs are 0.781 (on-axis) and 0.853 (off-axis). These results confirm the reliability of our simulator.

4.6. Applications

Fizeau Interferometer: We simulate Fizeau interferometers [15] as follows: a coherent input wavefront (650 nm) reflects off the test surface, whose profile determines the interference patterns captured by the sensor. Because of the coherency of the light source, all waves should interfere at the sensor plane and hence Eq. (4) becomes a coherent summation of all wavelets from all scene points. The reference measurement is generated with a surface parameterized by reference curvature and quadratic coefficients, as shown in Fig. 10 (c) and (d). We then employ differentiable rendering to recover the surface parameters, initialized with randomly perturbed values (Fig. 10 (d)) with corresponding measurement (Fig. 10 (a)). The optimization is driven by the MSE between the recovered and reference measurements. Because our wave optics model accurately captures phase interference, which reflects surface structures, both the surface (Fig. 10 (d)) and measurement (Fig. 10 (b)) are accurately recovered. This experiment demonstrates the applicability of the proposed model to coherent interference.

Freeform Optics: We perform differentiable rendering on freeform optics imaging, which is obtained by illuminating the surface with a coherent plane wave (wavelength: 650 nm) and accounting for coherent ray interactions. Specifically, we recover the target measurement in Fig. 11 (d) by surface optimization. The surface is randomly initialized with measurement in Fig. 11 (a), and we conduct ray- and wave-optimization for surface recovery, both are guided by minimizing the MSE between rendered and target measurements. Because of its coherent nature, accounting for wave optics is required for accurate light propagation. Therefore, as shown in Fig. 11 (b) and (c), the recovery is accurate only when wave optics effects are incorporated. These results underscore the versatility and importance of our differentiable wave optics simulator in non-lens optical systems.

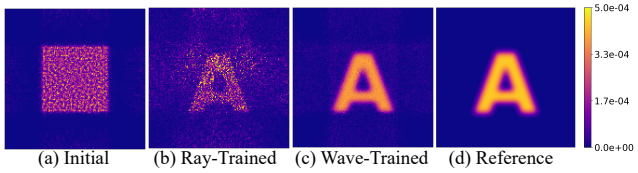


Figure 11. **Optimizing a freeform optical surface under coherent illumination**. Setup: A monochromatic plane wave is modulated by a freeform optical surface. Due to its coherence, the modulated wavefront interferes with itself in propagation, which can only be accounted for by wave optics. As a result, the wave-trained surface yields accurate recovery, which is not achievable by the ray-trained one. Sensor size: 5.8 mm^2.

4.7. Limitations

Although our simulator outperforms existing methods, its fidelity and efficiency remain constrained by system scale and aberration strength. For example, large-aperture systems with strong aberrations require very high sampling rates to accurately model wavefronts and propagation [16]. Similar issues arise in systems lacking a well-defined focal length or aperture stop, where the wavefront deviates significantly from a nominal sphere. In such cases, using a reference sphere can lead to large residual errors and high sampling demands. Aligning the wavefront by constant OPL may improve sampling efficiency, and exploring such alternative reference geometries is left for future work.

5. Conclusion

End-to-end optimization leverages the interplay between optics and computational algorithms, but existing frameworks lack the accuracy and efficiency to assess wave optics requirements. We present an efficient, differentiable wave optics simulator that reveals how diffraction impacts joint lens and algorithm design. Experiments show that neglecting diffraction leads to suboptimal configurations and degraded performance under diffraction-limited conditions. These results underscore the need for physics-aware modeling, further validated through differentiable rendering for Fizeau interferometers and freeform optics.

Acknowledgements

This work was supported in part by the Early Career Faculty Development Award for N. Antipa and T.-M. Li from the Jacobs School of Engineering at UC San Diego, the Ronald L. Graham Chair and the UC San Diego Center for Visual Computing. We also acknowledge NSF grant 2341952, ONR grant N00014-23-1-2526, and gifts from Adobe, Google, Qualcomm and Rembrand.

References

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 126–135, 2017. 5

[2] Seung-Hwan Baek, Hayato Ikoma, Daniel S Jeon, Yuqi Li, Wolfgang Heidrich, Gordon Wetzstein, and Min H Kim. Single-shot hyperspectral-depth imaging with learned diffractive optics. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2651–2660, 2021. 1, 2

[3] Hossein Baktash, Yash Belhe, Matteo Giuseppe Scopelliti, Yi Hua, Aswin C Sankaranarayanan, and Maysamreza Chamanzar. Computational imaging using ultrasonically-sculpted virtual lenses. In *2022 IEEE International Conference on Computational Photography (ICCP)*, pages 1–12. IEEE, 2022. 3

[4] Shiqi Chen, Huajun Feng, Dexin Pan, Zhihai Xu, Qi Li, and Yueting Chen. Optical aberrations correction in postprocessing using imaging simulation. *ACM Transactions on Graphics (TOG)*, 40(5):1–15, 2021. 1, 2, 5, 6

[5] Geoffroi Côté, Fahim Mannan, Simon Thibault, Jean-François Lalonde, and Felix Heide. The differentiable lens: Compound lens search over glass surfaces and materials for object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20803–20812, 2023. 1, 2

[6] Alice Fontbonne, Hervé Sauer, and François Goudail. Comparison of methods for end-to-end co-optimization of optical systems and image processing with commercial lens design software. *Optics Express*, 30(8):13556–13571, 2022. 2

[7] Joseph W Goodman. *Introduction to Fourier optics*. Roberts and Company publishers, 2005. 1, 2, 4, 5

[8] Aymeric Halé, Pauline Trouvé-Peloux, and J-B Volatier. End-to-end sensor and neural network design using differential ray tracing. *Optics express*, 29(21):34748–34761, 2021. 2

[9] Tianyue He, Qican Zhang, Chongyang Zhang, Tingdong Kou, and Junfei Shen. Learned digital lens enabled single optics achromatic imaging. *Optics Letters*, 48(3):831–834, 2023. 1, 2

[10] Francis Arthur Jenkins and Harvey Elliott White. Fundamentals of optics. *Indian Journal of Physics*, 25:265–266, 1957. 3

[11] Michael R Kellman, Emrah Bostan, Nicole A Repina, and Laura Waller. Physics-based learned design: optimized coded-illumination for quantitative phase imaging. *IEEE Transactions on Computational Imaging*, 5(3):344–353, 2019. 2

[12] Diederik P Kingma. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 5

[13] Zongling Li, Qingyu Hou, Zhipeng Wang, Fanjiao Tan, Jin Liu, and Wei Zhang. End-to-end learned single lens design using fast differentiable ray tracing. *Optics Letters*, 46(21):5453–5456, 2021. 2, 3

[14] Yuankun Liu, Chongyang Zhang, Tingdong Kou, Yueyang Li, and Junfei Shen. End-to-end computational optics with a singlet lens for large depth-of-field imaging. *Optics express*, 29(18):28530–28548, 2021. 2

[15] Daniel Malacara. Optical shop testing. *Wiely Interscience*, 2007. 8

[16] Zemax Manual. Optical design program. 2011. 2, 3, 4, 5, 6, 8

[17] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. 2017. 4, 6

[18] Yifan Peng, Qilin Sun, Xiong Dun, Gordon Wetzstein, Wolfgang Heidrich, and Felix Heide. Learned large field-of-view imaging with thin-plate optics. *ACM Trans. Graph.*, 38(6):219–1, 2019. 1, 2

[19] Stanislav Pidhorskyi, Timur Bagautdinov, Shugao Ma, Jason Saragih, Gabriel Schwartz, Yaser Sheikh, and Tomas Simon. Depth of field aware differentiable rendering. *ACM Transactions on Graphics (TOG)*, 41(6):1–18, 2022. 2

[20] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015. 5

[21] Robert R Shannon. *The art and science of optical design*. Cambridge University Press, 1997. 3, 4

[22] Zheng Shi, Yuval Bahat, Seung-Hwan Baek, Qiang Fu, Hadi Amata, Xiao Li, Praneeth Chakravarthula, Wolfgang Heidrich, and Felix Heide. Seeing through obstructions with diffractive cloaking. *ACM Transactions on Graphics (TOG)*, 41(4):1–15, 2022. 1, 2

[23] Vincent Sitzmann, Steven Diamond, Yifan Peng, Xiong Dun, Stephen Boyd, Wolfgang Heidrich, Felix Heide, and Gordon Wetzstein. 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):1–13, 2018. 1, 2

[24] Qilin Sun, Congli Wang, Fu Qiang, Dun Xiong, and Heidrich Wolfgang. End-to-end complex lens design with differentiable ray tracing. *ACM Trans. Graph*, 40(4):1–13, 2021. 1, 2, 3

[25] Shiyu Tan, Yicheng Wu, Shoou-I Yu, and Ashok Veeraraghavan. Codedstereo: Learned phase masks for large depth-of-field stereo. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7170–7179, 2021. 1

[26] Ethan Tseng, Ali Mosleh, Fahim Mannan, Karl St-Arnaud, Avinash Sharma, Yifan Peng, Alexander Braun, Derek Nowrouzezahrai, Jean-Francois Lalonde, and Felix Heide. Differentiable compound optics and processing pipeline optimization for end-to-end camera design. *ACM Transactions on Graphics (TOG)*, 40(2):1–19, 2021. 2

[27] Congli Wang, Ni Chen, and Wolfgang Heidrich. do: A differentiable engine for deep lens design of computational imaging systems. *IEEE Transactions on Computational Imaging*, 8:905–916, 2022. 2, 3, 4

[28] Zongzhao Wang, Olga Baladron-Zorita, Christian Hellmann, and Frank Wyrowski. Generalized debye integral. *Optics Express*, 28(17):24459–24470, 2020. 2

[29] Haoyu Wei, Xin Liu, Xiang Hao, Edmund Y Lam, and Yifan Peng. Modeling off-axis diffraction with the least-sampling angular spectrum method. *Optica*, 10(7):959–962, 2023. 1, 2, 3, 5, 6

[30] Frank Wyrowski and Michael Kuhn. Introduction to field tracing. *Journal of Modern Optics*, 58(5-6):449–466, 2011. 2

[31] Xinge Yang, Qiang Fu, Mohamed Elhoseiny, and Wolfgang Heidrich. Aberration-aware depth-from-focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023. 1, 3

[32] Xinge Yang, Qiang Fu, Yunfeng Nie, and Wolfgang Heidrich. Image quality is not all you want: Task-driven lens design for image classification. *arXiv preprint arXiv:2305.17185*, 2023. 1, 2

[33] Xinge Yang, Matheus Souza, Kunyi Wang, Praneeth Chakravarthula, Qiang Fu, and Wolfgang Heidrich. End-to-end hybrid refractive-diffractive lens design with differentiable ray-wave model. In *SIGGRAPH Asia 2024 Conference Papers*, pages 1–11, 2024. 1, 2, 3, 5, 6

[34] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018. 5

[35] Rongshuai Zhang, Fanjiao Tan, Qingyu Hou, Zongling Li, Zaiwu Sun, Changjian Yang, and Xiangyang Gao. End-to-end learned single lens design using improved wiener deconvolution. *Optics Letters*, 48(3):522–525, 2023. 2