

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, rramamoorthi, 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 in 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.

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, 29, 30]. Given a dataset of training images, the differentiable optics simulator models corresponding measurements taken by the optics system, and the computer vision algorithm extracts information from simulated measurements. 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

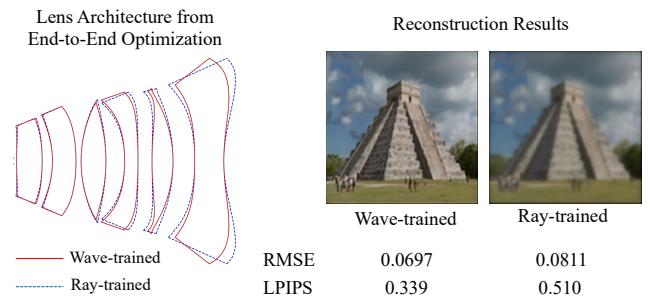


Figure 1. **End-to-end optimized lens architectures and reconstruction models using ray and wave optics.** By taking diffraction into account, our wave-trained model yields sharper reconstruction results than the baseline using ray optics.

overall system optimization. Computational expense imposes a limitation on exploring the effects of realistic optical modeling, particularly when integrating wave optics. To model the wave optics effects of a ray, we must account for its diffraction effects across the entire sensor rather than simply tracing its path. This complexity intensifies as the number of rays and sensor pixels increases. Therefore, many end-to-end designs simplify the physics models by using ray optics, which neglects diffraction, to simulate light transport [5, 24, 30]. Even though some simulators do take diffraction into account, they still make assumptions of thin-phase surfaces [22, 28] or shift-invariance [2, 9, 23]. These approaches fail to model realistic multi-element and compound optics designs, limiting the space of possible lens designs. Although recent frameworks model more realistic wave optics [4, 31], their accuracy and efficiency in different configurations remain questionable, and the significance of wave optic effects on system optimization remains an open problem.

In this paper, we propose an accurate, efficient, and differentiable optics simulator, which uses ray tracing with Rayleigh-Sommerfield integral [7] to model diffraction and off-axis aberrations in compound optical systems without thin-phase or paraxial approximations. To address the computation costs of modeling 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 solution with weaker diffraction. 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 optimization, and demonstrate that 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 designs construct a merit function, which combines lens properties and transfer function quality, to optimize optics systems [16]. Nonetheless, the merit functions are not guaranteed to reflect computer vision task performance [6, 30]. To address this issue, end-to-end optimizations use task performance to concurrently improve hardware and software. Through the cooperation between differentiable optics simulators and inference algorithms on a large dataset [5, 9, 24, 30], end-to-end optimization provides a data-driven design that addresses the interdependence among optics, algorithms, and tasks [6].

End-to-end optimization has been widely applied to image reconstruction [6, 13, 14, 18, 22, 23] and restoration [8, 33]. Sitzmann *et al.* extend the depth of field on computational cameras [23]. Peng *et al.* achieve high FoV image reconstruction [18]. Shi *et al.* recover unobstructed scenes by a diffractive optical element and a point-PSF-aware neural network [22]. The strategy has also been applied to semantic information extraction. Baek *et al.* acquire depth information from hyperspectral imaging by jointly optimizing diffractive optical elements and a network [2]. Kellman *et al.* recover phase information by jointly optimizing coded-illumination patterns for an LED array and an unrolled physics-based network [11]. Pidhorskyi *et al.* develop a differentiable ray tracer for depth-of-field aware

scene intensity recovery [19]. Yang *et al.* optimize off-axis aberration performance for image classification [30]. Cote *et al.* optimize lens materials and structures for object detection [5]. In these frameworks, end-to-end optimization offers task-specific optics and algorithms.

2.2. Balancing Accuracy and Efficiency in Differentiable Optics Simulation

A notable concern in end-to-end optimization is the computational cost of differentiable optics simulation [27, 31]. Because end-to-end design usually requires complex physics simulation to generate measurements from a large imaging dataset, complicated gradient propagations are needed to model the relation among optics parameters, measurements, and semantic information [27]. To address the cost in large FoV differentiable rendering, simplified physics models are usually adopted. Peng *et al.* adopt the thin phase assumption to optimize thin-plate lens in large FoV imaging reconstruction [18]. Sun *et al.* use ray optics to model differentiable ray tracing in complex lens model [24]. Cote *et al.* use ray optics with ray aiming to improve the accuracy of simulating lenses with strong pupil aberrations [5]. Although these methods manage to reduce the computation costs, their approaches cannot model wave optical effects.

It is also common to model simplified wave optical effects. Sitzmann *et al.* use Fresnel propagation to model wave optics in diffractive optics but assume the system is shift-invariant [23]. Shi *et al.* incorporates diffractive optics and a lens using the thin phase assumption [22]. He *et al.* compute PSFs with diffraction theory in shift-variant systems [9]. Tseng *et al.* replace the entire imaging pipeline with a neural network to render PSFs [26]. Wei *et al.* model off-axis diffraction using the angular spectrum method (ASM) [28], but assume the system is a thin-phase single lens. All these assumptions limit their applicability to compound optical systems.

Chen *et al.* [4] developed a ray-wave optics simulator that calculates diffraction at the exit pupil. A concurrent work by Yang *et al.* uses an ASM-based approach to model a hybrid lens-diffractive optic system [31]. Although these two frameworks provide more realistic wave optics modeling, their efficiency and robustness to various lens systems remain questionable. Moreover, an unexplored problem for these frameworks is the difference between using ray and wave models during lens and network training. In our work, we adopt a differentiable rendering method that models diffraction from rays and use it to explore the significance of accurate wave modeling in end-to-end design.

3. Differentiable Optics Model

Our differentiable simulator is designed to accurately and efficiently capture both aberration and diffraction in com-

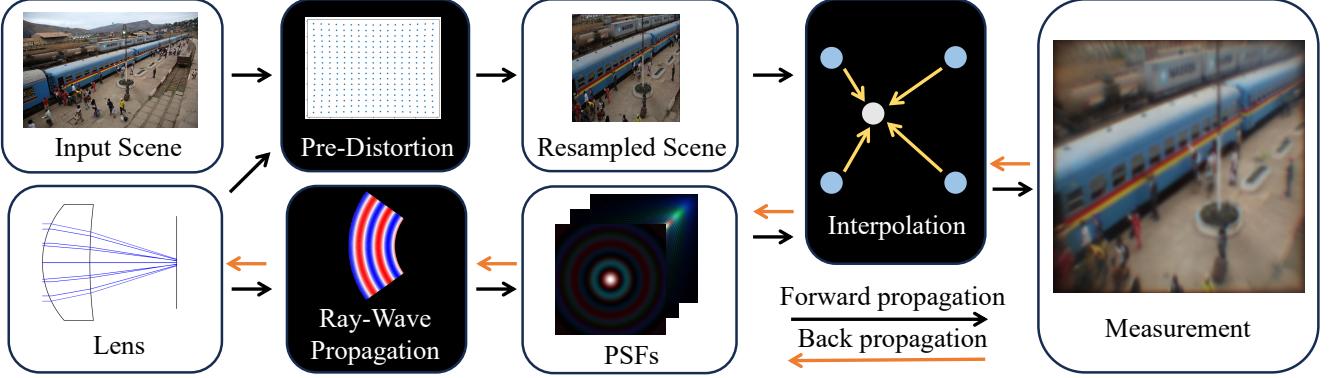


Figure 2. **Our proposed differentiable wave optics simulator.** Given an input scene and lens configuration, we first resample the scene based on the lens’ pre-distortion map. Next, we generate diffraction-aware PSFs using our wave optics simulator. Finally, we interpolate the convolution of the resampled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

ound 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]:

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

However, directly computing Eq. 1 across the entire FoV is computationally intensive, requiring full-resolution PSF rendering for every point source. To address this challenge, 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 corresponding sub-scene intensities [3]. Details of this interpolation technique are provided in Sec. 3.2. With accurate and efficient optical modeling, including diffraction, off-axis aberrations, and geometric distortions, the simulator improves the robustness of data-driven lens design and allows for a deep exploration of wave optics effects in imaging applications.

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

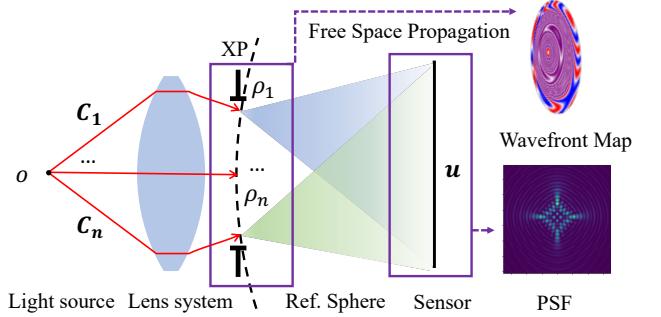


Figure 3. **Our wave optics simulator.** We trace rays emitted from a point source o to the reference sphere on the system’s exit pupil, and compute intersections $\{\rho_i\}$ and associated phase on a wavefront map. We then perform free-space propagation toward the sensor to generate a PSF. XP: Exit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

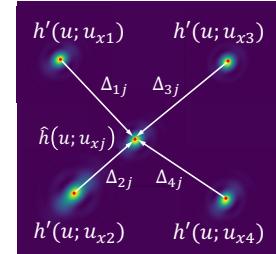


Figure 4. **Approximating unsampled PSFs.** Our system first samples PSFs $h'(\mathbf{u}; \cdot)$ on a regular grid. Next, by exploiting the isoplanatic property, it approximates off-grid PSFs $\hat{h}(\mathbf{u}; \cdot)$ by interpolating shifted and scaled versions of nearest samples $h'(\mathbf{u};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

[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 chief ray and the sensor plane, and the distance between the exit pupil and the sensor, respectively.

Therefore, the problem amounts to calculating the ampli-

tude and phase of the complex field at the reference sphere. By noting that the exit pupil is an image of the aperture stop, we model the amplitude by the square root of the aperture stop transmittance. On the other hand, the phase at the reference sphere is determined by the optical path length (OPL) δ calculated by

$$\delta = \int_C n(s) ds, \quad (2)$$

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-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 [28, 29], the spherical structure effectively reduces the phase variation, and hence alleviates the sampling requirement. In other words, we model the wavefront error map on the reference sphere 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 $\rho_i = (\rho_{x_i}, \rho_{y_i}, \rho_{z_i})$, we model the complex field by

$$v(\rho_i) = a_i \exp(jk\delta_i), \quad (3)$$

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

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

where $\vec{r}_{u,i}$ denotes the vector from ρ_i to sensor coordinate \mathbf{u} , and $\theta_{u,i}$ is the angle between $\vec{r}_{u,i}$ and normal vector of the reference geometry at ρ_i .

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 chief ray emitting from \mathbf{x} and the sensor plane. Because the function is one-to-one, the scene intensities $b(\mathbf{x})$ and PSF $h(\mathbf{u}; \mathbf{x})$ 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:

$$\hat{h}(\mathbf{u}; \mathbf{u}_{xj}) = \sum_i w_i(\mathbf{u}_{xj}) h'(\mathbf{u} - \Delta_{ij}; \mathbf{u}_{xi}), \quad (5)$$

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:

$$\begin{aligned} I(\mathbf{u}) &= \sum_{\mathbf{u}_x} b'(\mathbf{u}_x) \sum_i w_i(\mathbf{u}_x) h'(\mathbf{u} + \mathbf{u}_{xi} - \mathbf{u}_x; \mathbf{u}_{xi}) \\ &= \sum_i \sum_{\mathbf{u}_x} b''_i(\mathbf{u}_x) h'(\mathbf{u} + \mathbf{u}_{xi} - \mathbf{u}_x; \mathbf{u}_{xi}), \end{aligned} \quad (6)$$

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:

$$\begin{aligned} I(\mathbf{u}) &= \sum_i \sum_{\mathbf{u}_x} b''_i(\mathbf{u}_x) h_i(\mathbf{u} - \mathbf{u}_x) \\ &= \sum_i b''_i * h_i \end{aligned} \quad (7)$$

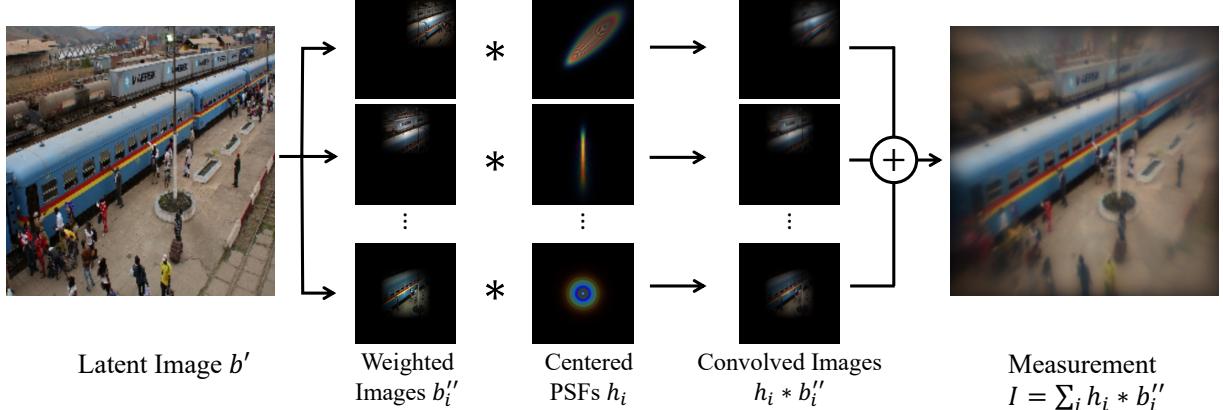


Figure 5. **Rendering measurement with a subset of PSFs.** Given a latent image b' , we first generate weighted images b_i'' . Next, we generate PSFs h_i at the centers of weighted images and pair them to corresponding PSFs. Finally, we convolve weighted images and PSFs ($h_i * b_i''$) and sum them up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

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.

Using the chain rule, gradients can be back-propagated from the measurements I through the wave-optics PSFs $h(\mathbf{u})$, the complex field $v(\rho_i)$ on the reference sphere, and ultimately to the lens parameters. This differentiability enables precise modeling and analysis of the interactions between lens configurations and wave-optics effects on the measurements. 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.

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 physics 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, 28, 31] under various conditions. Specifically, we render 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. The Huygens PSF computed using Zemax serves as the reference, and we quantify similarity to it using the structural similarity index (SSIM). Additionally, we evaluate the efficiency by the ray count and the computational time for each simulator.

Among ray-tracing-based approaches, our method demonstrates superior accuracy and efficiency across most scenarios. The only exception occurs in the in-focus case, where our simulator requires more rays than Zemax-FFT does. However, our approach achieves higher accuracy while maintaining a shorter runtime, highlighting its efficiency and precision.

It is notable that ASM-based methods, Wei et al. [28], Yang et al. [31], are sensitive to defocus. This is because as the system becomes increasingly defocused, phase variations across the pupil plane and propagation kernel are extremely rapid. However, ASM requires discretizing pupil field and propagation function on a 2D grid, which restricts the ability to capture rapid phase variation, resulting in degraded accuracy and efficiency for ASM-based approaches. Although Chen et al. [4] allows flexible ray distribution, their wave modeling does not account for the magnitude changes brought by $|\vec{r}_{u,i}|$ in Eq. 4, and use projected $|\vec{r}_{u,i}|$ onto ray directions, instead of actual values. Therefore, their method fails to capture the magnitude changes over large spot sizes caused by defocusing. Furthermore, defocusing also makes Airy disk, a common tool to evaluate diffraction in perfect lenses, an unreliable approach to model wave optics effects. The results show that our simulator provides higher robustness, accuracy, and efficiency, in defocusing and large FoV systems, which easily occur in end-to-end optimization.

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.

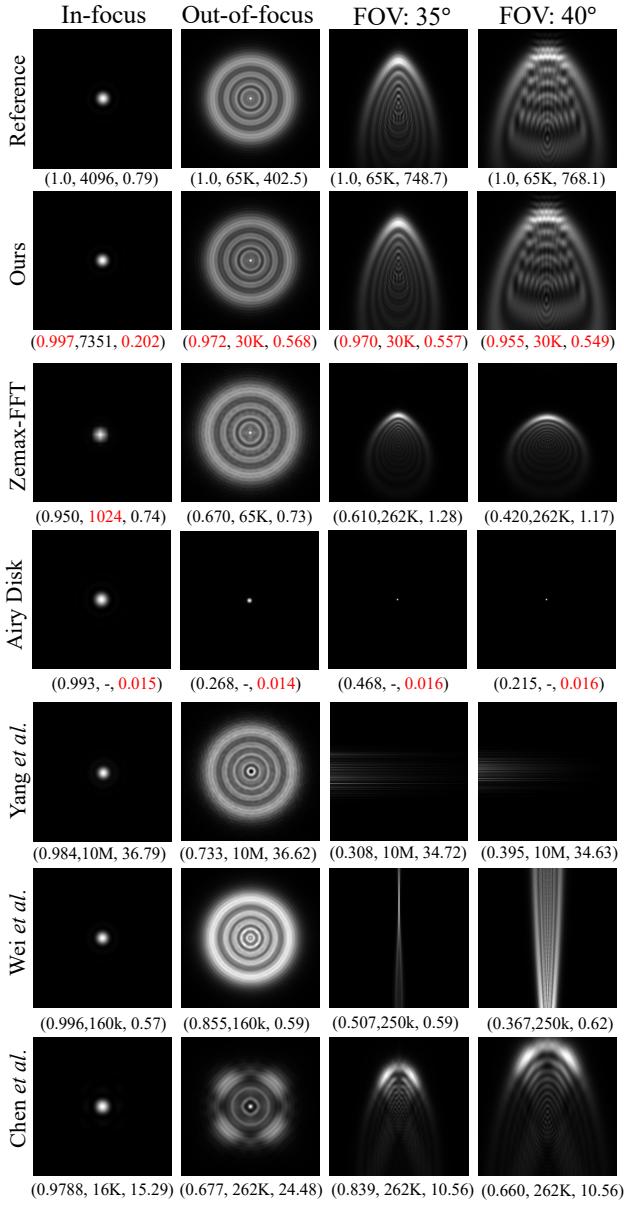


Figure 6. PSFs rendered by different simulators under different conditions. Unlike existing simulators [4, 16, 28, 29], ours avoids wavefront discretization and remains robust to defocus and large FoVs, achieving the highest accuracy and efficiency. The tuple (SSIM, ray count, time in sec.) highlights 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.

To compare their robustness to diffraction effects, we use wave optics in 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 is a weighted sum of root-mean-square error (RMSE)

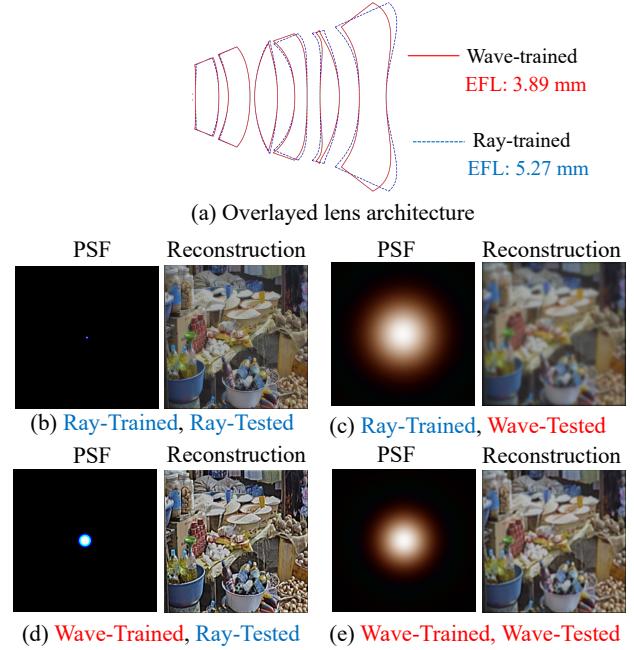


Figure 7. An analysis of ray- and wave-trained systems. The ray-trained system pursues minimal geometric spot size (b) but neglects diffraction blur (c). The wave-trained system has a larger geometric spot size (d), but a lower effective focal length (EFL) to control diffraction, yielding better reconstruction performance (e). PSF size: 0.044 mm²

and perceptual loss (LPIPS) [32] between the normalized input scene intensities and the reconstructed results. To keep a consistent FoV for fair comparisons, whenever the focal length varies in optimization, 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 (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 consistently uses wave optics in evaluation. 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), the wave-trained lens changes its architecture to shorten the focal length and weaken diffraction. On the other hand, the ray-trained lens focuses on minimizing RMS spot size, as shown in Fig. 7 (b).

Although the ray-trained system achieves a smaller geometric spot size, as shown in Fig. 7 (b) and (d), it fails to ac-

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)

count 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. 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 singlet lens at a 30° off-axis field point with wavelength 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 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 assess interpolation accuracy across different FoVs and lens complexities, we use measurements rendered with 969

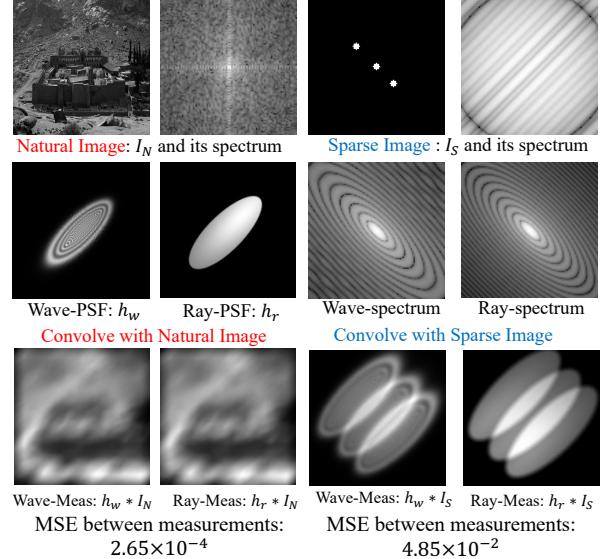
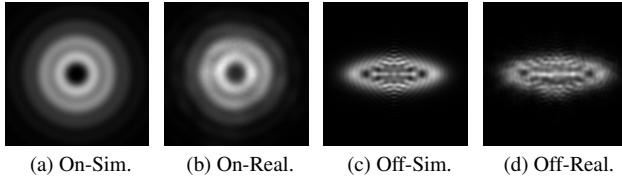


Figure 8. **Comparing ray and wave measurements in an aberration limited system.** The key spectral difference between ray- and wave-PSFs lies in high frequencies, affecting measurements only when the image has rich high-frequency components. Thus, for natural images, both systems receive similar training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

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	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
Triplet	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
Six	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
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

PSFs, the maximum feasible under hardware constraints, as a reference and compute its disparity to sparser interpolated measurements, using PSF counts ranging from 9 to 289. As summarized in Table 2, systems with larger FoVs require more PSFs to reduce the discrepancy between interpolated and reference measurements. This is due to increased aberrations in wide FoVs, which weaken the isotropicity among PSFs. Furthermore, as lens complexity increases, the system has weaker aberration, and fewer PSFs are needed for accurate rendering. Table 2 also reports the computational time of interpolating a single image. As shown, with denser interpolation, end-to-end optimization



(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 .

becomes computationally expensive. Thus, selecting an appropriate number of PSFs is crucial to balancing computational efficiency and rendering accuracy, with the optimal choice depending on the FoV and lens complexity.

4.5. Hardware Validation

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

4.6. Applications

Fizeau Interferometer We apply the simulator to a Fizeau interferometer [15] as follows: a coherent input wavefront (wavelength: 650 nm) reflects off the test surface, whose profile determines the interference patterns captured by the sensor. To establish a reference, we first generate measurements with a reference surface parameterized by 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

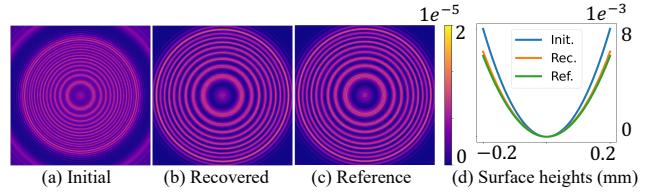


Figure 10. Recovering a quadratic surface based on Fizeau interferometer measurements. Setup: A coherent wavefront is reflected by a quadratic surface, and the resulting interference pattern is detected by the sensor. The interference pattern is determined by the surface geometry. By accurately modeling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

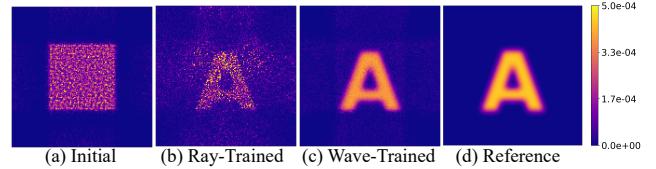


Figure 11. Optimizing a freeform optical surface under coherent illumination. Setup: A monochromatic plane wave is modulated by a freeform optical surface. Due to its coherence, the modulated wavefront interferes with itself in propagation, which can only be accounted for by wave optics. As a result,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

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.

5. Conclusion

End-to-end optimization exploits the interdependence between optics and computational algorithms in imaging systems. However, due to insufficient accuracy and efficiency, existing frameworks have not analyzed the requirements of modeling wave optics in simulation in system designs. In this paper, we develop an efficient, accurate, and differentiable wave optics simulator to analyze the role of diffraction in end-to-end optimization. Experimental results show that modeling light transport with and without diffraction yields different lens configurations and algorithm adaptations. When diffraction is not considered in system design, performance degradation can occur under diffraction-limited conditions. These findings highlight the critical importance of physics-aware modeling for imaging system designs. The importance of modeling wave optics is further demonstrated in differentiable rendering in Fizeau interferometer and freeform optics.

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. 6
- [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 Yuetong Chen. Optical aberrations correction in postprocessing using imaging simulation. *ACM Transactions on Graphics (TOG)*, 40(5):1–15, 2021. 1, 2, 3, 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, 3, 4
- [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. 4
- [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. 6
- [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
- [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. *Wiley Interscience*, 2007. 8
- [16] Zemax Manual. Optical design program. 2011. 2, 4, 5, 6
- [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. 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. 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, Shou-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
- [28] 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](#), [4](#), [5](#), [6](#)
- [29] Xinge Yang, Qiang Fu, Mohamed Elhoseiny, and Wolfgang Heidrich. Aberration-aware depth-from-focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023. [1](#), [4](#), [6](#)
- [30] 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](#)
- [31] 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](#), [5](#)
- [32] 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. [6](#)
- [33] 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](#)