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

Supplementary Material

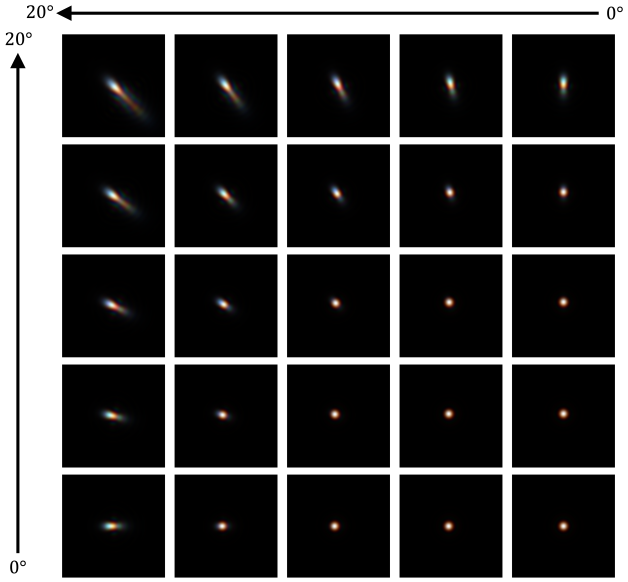


Figure 1. **On-axis and off-axis PSFs rendered from a Cooke Triplet lens.** We display PSFs generated by the incident angle from 0 to 20 degrees.

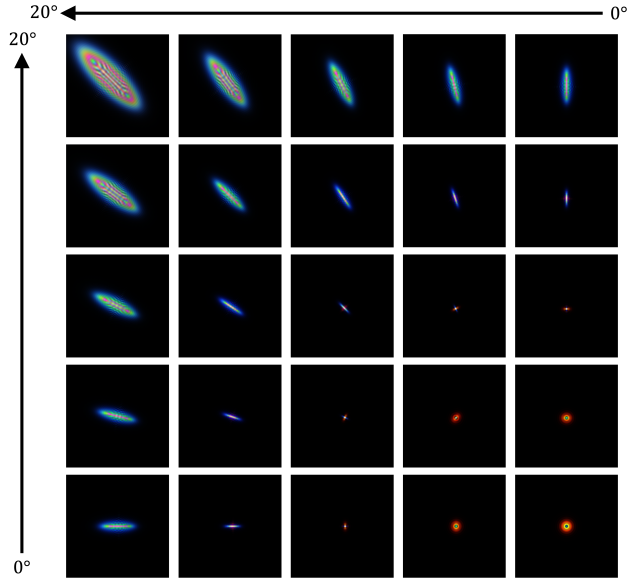


Figure 2. **On-axis and off-axis PSFs rendered from a singlet lens.** We display PSFs generated by the incident angle from 0 to 20 degrees.

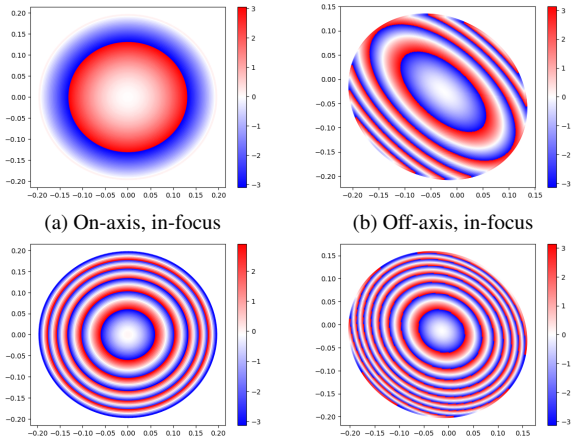
In this supplemental document, we present additional results of the proposed differentiable wave optics simulator to support our findings.

1. Wave Optics Rendering

In this section, we first demonstrate point spread functions (PSFs) and wavefront maps rendered from different incident angles, apertures, and lens configurations. We also show the measurement rendered by interpolating wave-rendered PSFs and analyze the impact of interpolation. Moreover, we analyze the computation efficiency under different physics models.

1.1. PSFs across Different FoVs

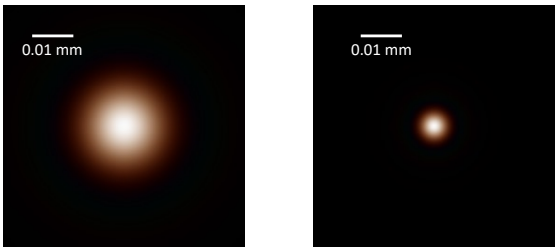
Fig. 1 and 2 illustrate the PSFs rendered across incident angles from 0° to 20° for Cooke Triplet and singlet lenses, respectively. The results demonstrate that our simulator effectively captures variations in PSFs due to changing incident angles. When increasing the incident angle, more astigmatism and coma effects appear on the PSFs. Notably, compared with the PSFs of the Cooke Triplet lens, off-axis aberrations, and coma are significantly more pronounced in the PSFs of the singlet lens. This result highlights the difference between the aberration degree of these two types of lenses.



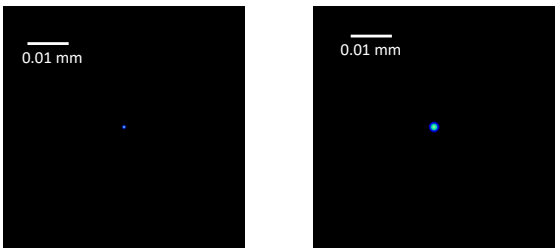
(a) On-axis, in-focus (b) Off-axis, in-focus

(c) On-axis, out-of-focus (d) Off-axis, out-of-focus

Figure 3. **The phase variations on the wavefront maps under different situations.** As the system becomes off-axis or out-of-focus, the phase changes more rapidly on the wavefront map. Unit: mm.



(a) Aperture radius: 0.1 mm, wave-PSF (b) Aperture radius: 0.3 mm, wave-PSF



(c) Aperture radius: 0.1 mm, ray-PSF (d) Aperture radius: 0.3 mm, ray-PSF

Figure 4. **PSFs rendered with different aperture radii and physics models.** With a larger aperture radius (0.3 mm), the spot size of wave-PSF and Ray-PSF becomes closer to each other.

1.2. Wavefront Maps

We present wavefront maps for a singlet lens under various conditions in Fig. 3. Fig. 3a shows that the wavefront map generated at on-axis and in-focus configuration has the slightest phase variation. In contrast, Fig. 3b shows that off-axis rays introduce significant aberrations, leading to phase shifts spanning over five wavelengths and causing defocus blur in the corresponding PSF. Additionally, by shifting the sensor, we make the system out-of-focus and display PSFs in Fig. 3c and 3d. As observed, the phase variations in the wavefront maps are more pronounced for both on-axis and off-axis beams, highlighting the degree of defocusing. These findings validate that the phase variation on the wavefront map reflects focusing conditions. By accurately modeling phase variations, our simulator is capable of modeling aberration and diffraction in various lens configurations.

1.3. PSFs Rendered under Different Aperture Sizes

Figure 4 illustrates how increasing the aperture radius reduces the diffraction spot size. When the aperture radius is 0.1 mm, because the system is diffraction limited, the spot size of wave-rendered PSF is much greater than that of ray-rendered PSF. As the aperture radius increases to 0.3 mm, the diffraction becomes weaker, and geometric spot size increases, bringing the diffraction and aberration limits closer. These findings confirm that the disparity between ray-trained and wave-trained systems depends on aperture size: a larger aperture radius diminishes the differences between PSFs rendered by ray optics and wave optics, indicating reduced sensitivity of system optimization to diffraction effects.

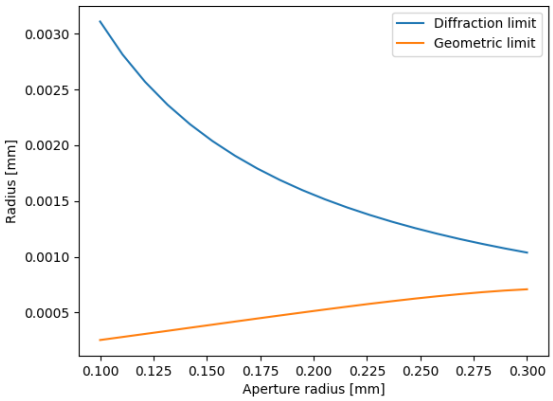
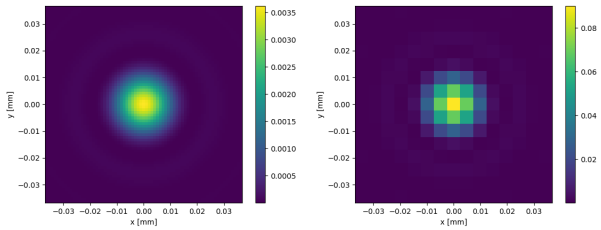


Figure 5. **The diffraction and aberration (geometric) limits under different aperture radii.** As the aperture radius increases, the two limits gradually converge into the same value. The values are measured from a Cooke triplet lens.



(a) Pixel width: 1um. (b) Pixel width: 5 um.

Figure 6. **PSFs rendered with different pixel widths.** The diffraction effects become less obvious when using a coarser resolution.

Figure 5 illustrates the change of diffraction and geometric limits with aperture radius. For an aperture radius of 0.1 mm, the diffraction limit significantly exceeds the aberration limit. Therefore, the system has enough flexibility to adjust its configuration, sacrificing aberration and achieving better diffraction performance. However, as the aperture radius increases, the diffraction and geometric limits converge, reducing the potential gains from this trade-off. Consequently, systems with larger apertures exhibit diminished flexibility in optimizing between diffraction and aberration performance.

1.4. PSFs with Different Pixel Sizes

We vary pixel size to investigate the dependency of lens design solutions on pixel area and summarize results in Table 1. As observed, when the pixel size is 1 μm, ray-trained and wave-trained systems have a large disparity. However,

Table 1. Reconstruction performance on wave optics rendered measurements (RMSE / LPIPS)

PS	Training physics		MF	CL
	Wave	Ray		
Aperture radius: 0.1 mm				
1	**0.106 / 0.265**	0.148 / 0.772	8.689	0.869
5	**0.099 / 0.176**	0.107 / 0.437	1.906	0.905
10	**0.071 / 0.060**	0.074 / 0.116	0.229	0.999
Aperture radius: 0.3 mm				
1	**0.104 / 0.230**	0.112 / 0.483	0.015	0.999
5	**0.069 / 0.064**	0.071 / 0.065	0.303	0.999
10	**0.062 / 0.059**	0.068 / 0.064	0.132	0.999

PS: pixel size (unit: µm), MF: Mismatch between F-numbers

CL: Cosine similarity between normalized and vectorized lens curvatures

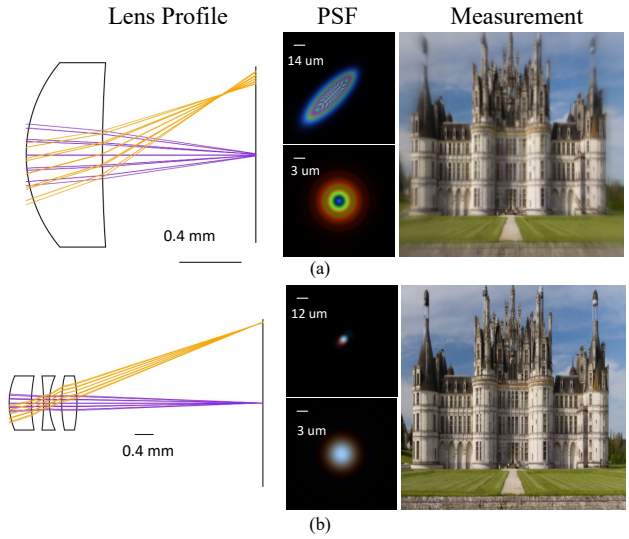


Figure 7. **Examples of rendered PSFs and measurement.** Our simulator accurately captures aberration and diffraction in on-axis and off-axis PSFs from (a) singlet lens and (b) Cooke triplet lenses.

as the pixel size increases, the disparity diminishes. Notably, we evaluate the mismatch between lenses by vectorizing their curvatures and measure the cosine similarity, abbreviated as CL in Table 1.

As visualized in Fig. 6, when increasing the pixel width from 1 to 5 µm, the PSF structures become coarser, with diffraction effects appearing less pronounced due to reduced pixel resolution. This observation aligns with our findings that smaller pixel sizes enhance the sensitivity of system optimization to wave optics, underscoring the importance of wave optics modeling for achieving high physical accuracy in imaging systems with high pixel resolutions.

Table 2. The expected and reference reconstruction performance of system trained by a subset of PSFs (RMSE / LPIPS)

N_T	Expected	Reference	→ (s)	← (s)
Singlet Lens				
9	0.069 / 0.138	0.098 / 0.226	2.21	7.46
25	0.077 / 0.125	0.082 / 0.176	3.94	17.14
81	0.079 / 0.154	0.082 / 0.159	11.09	52.39
289	0.084 / 0.164	0.084 / 0.164	31.26	176.72
969	0.082 / 0.149	0.082 / 0.149	113.55	679.97
Cooke Triplet Lens				
9	0.106 / 0.264	0.106 / 0.265	3.08	10.18
25	0.101 / 0.240	0.101 / 0.241	5.92	23.81
81	0.101 / 0.240	0.101 / 0.241	16.51	72.41
289	0.102 / 0.242	0.102 / 0.243	55.53	253.7
969	0.089 / 0.234	0.089 / 0.234	198.13	975.2

N_T: The number of PSFs used in approximation and optimization

→ (←): Time elapsed in the forward (backward) propagation

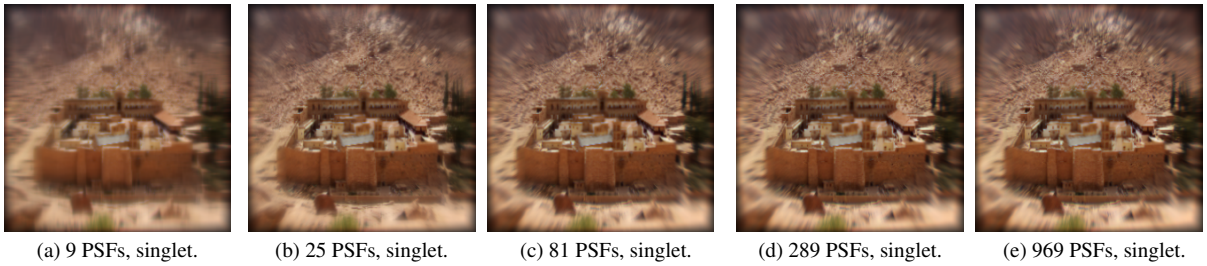
1.5. Measurement Rendering

We simulate PSFs and measurements obtained from singlet and Cooke triplet lenses in Fig. 7. The singlet lens demonstrates prominent effects from both diffraction and off-axis aberrations. The on-axis PSF shows clear diffraction fringes, while the off-axis shows a combination of coma and astigmatism with clear diffraction effects. The Cooke triplet lens, by nature of its optimized design, drastically reduces aberrations. These results demonstrate how our model faithfully renders images for both aberration-limited and diffraction-limited systems, enabling end-to-end optimization of lenses across both regimes.

1.6. Interpolated Measurements

Figure 8 illustrates example measurements rendered with varying the number of PSFs in interpolation. For measurements rendered from the Cooke triplet lens, the rendering accuracy remains consistent regardless of the number of PSFs used in interpolation. In contrast, the rendering accuracy for the singlet lens strongly depends on the number of PSFs, with noticeable artifacts appearing when interpolation is performed using only 9 PSFs. These results emphasize the importance of selecting an adequate number of PSFs, particularly for lenses with limited aberration control.

In addition to computing the RMSE between sparsely and densely rendered PSFs, we also evaluate the robustness of systems optimized using a subset of PSFs. Particularly, we optimize the system using 9 to 969 PSFs, and measure expected and reference reconstruction performance defined as follows: Expected performance is evaluated on measurements rendered by the same number of PSFs used in training, and reference performance is evaluated on measurements rendered by 969 PSFs, the maximum achievable un-



(a) 9 PSFs, singlet. (b) 25 PSFs, singlet. (c) 81 PSFs, singlet. (d) 289 PSFs, singlet. (e) 969 PSFs, singlet.

(f) 9 PSFs, triplet. (g) 25 PSFs, triplet. (h) 81 PSFs, triplet. (i) 289 PSFs, triplet. (j) 969 PSFs, triplet.

Figure 8. **Measurements interpolated by different number of PSFs.** We generate measurements from 9 to 969 of interpolated PSFs in singlet and triplet lenses. As the number of PSFs increases, the measurements converge to the same appearance.

der hardware constraints.

As summarized in Table 2, optimizing a Cooke Triplet lens system with just 9 PSFs results in a narrow disparity between expected and reference reconstruction performance. Adding more PSFs further improves both metrics, but they remain closely aligned. In contrast, a singlet lens system optimized with only 9 PSFs tends to yield overly optimistic results, leading to significant degradation in reference performance. The disparity arises from the limited capacity of the singlet lens to correct off-axis aberrations, making systems trained with isoplanatic assumptions less robust. In comparison, due to a superior aberration control, the Cooke Triplet lens allows robust optimization even with a minimal PSF count. These findings underscore the sampling requirements for optimizing various lens types to balance efficiency and accuracy.

1.7. Computation Efficiency

We plot the time elapsed in forward (dash line) and backward (solid line) propagation under different physics models in Fig. 12. As we can see, due to the higher computational costs, wave optics take around twice and four times longer than ray optics in forward and backward propagation, respectively. The analysis demonstrates the need to alleviate the computational costs of driving back-propagation on differentiable wave optics.

2. Lens Architecture

In this section, we present further analysis of the lens architecture optimized under different physics models and the associated PSFs.

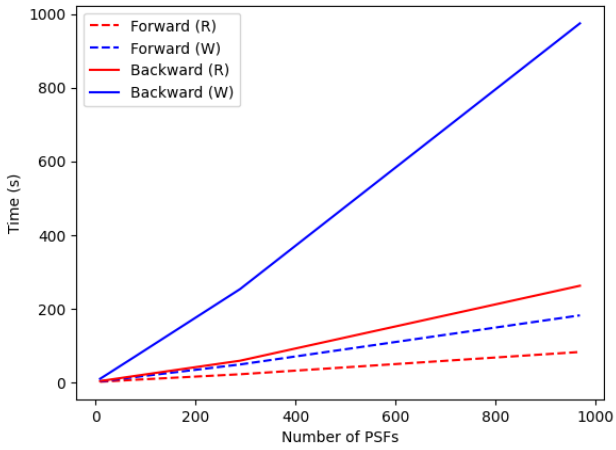


Figure 9. **Time elapsed in rendering.** We measure both forward (dash line) and backward (solid line) propagation using wave (W, blue) and ray (R, red) optics. Because of considering the interactions of every ray with every sensor grid, wave optics take longer than ray optics in both forward and backward propagation.

2.1. Change of F-Number in Optimization

To understand the divergence in optimization paths between the wave-trained and ray-trained lens, we plot the trend of the F-number of lenses over successive iterations in Fig. 10. During wave optics-based optimization, the system progressively reduces the F-number to control the PSF spot size. In contrast, under ray optics optimization, where the F-number has little influence on PSF structures, it remains close to the initial value throughout the training epochs.

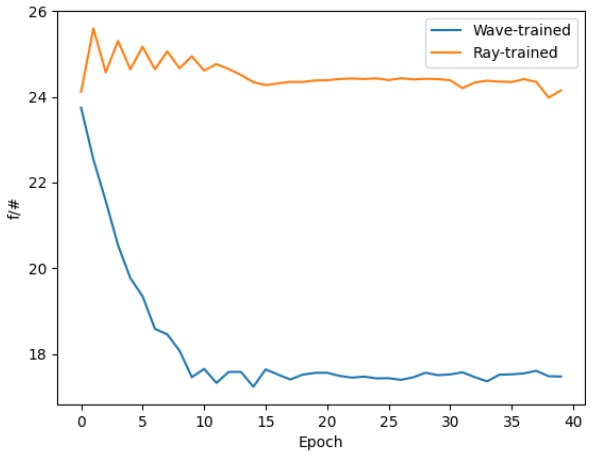
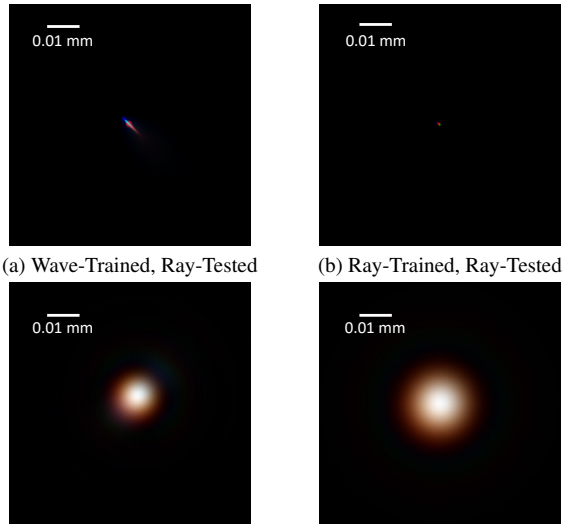


Figure 10. **The change of F-number in end-to-end optimization.** By pursuing a smaller diffraction spot size, the wave-trained system keeps reducing its F-number in optimization. On the other hand, the ray-trained system pursues a small geometric spot size, and the F-number remains around a constant.



(a) Wave-Trained, Ray-Tested (b) Ray-Trained, Ray-Tested

(c) Wave-Trained, Wave-Tested (d) Ray-Trained, Wave-Tested

Figure 11. **Off-axis PSFs of end-to-end optimized systems.** A wave-trained system sacrifices acceptable geometric RMS spot (a) size to achieve weaker diffraction (c). On the other hand, although a ray-trained system minimizes the geometric RMS spot size (b), it neglects the diffraction blur (d), which is the actual limiting factor in PSF.

2.2. Disparity between Off-Axis PSFs from Wave-Trained and Ray-Trained Systems

In Fig. 11, we compare off-axis PSFs generated by triplet lenses optimized with wave optics versus ray optics. The ray-optimized lens demonstrates superior control over geometric spot size. However, when wave optics effects are introduced during testing, the wave-optimized lens exhibits a superior ability to mitigate diffraction-related spot size. These results further underscore the importance of accounting for wave optics in end-to-end optimization, highlighting the overly optimistic performance estimates that can arise from neglecting diffraction effects.

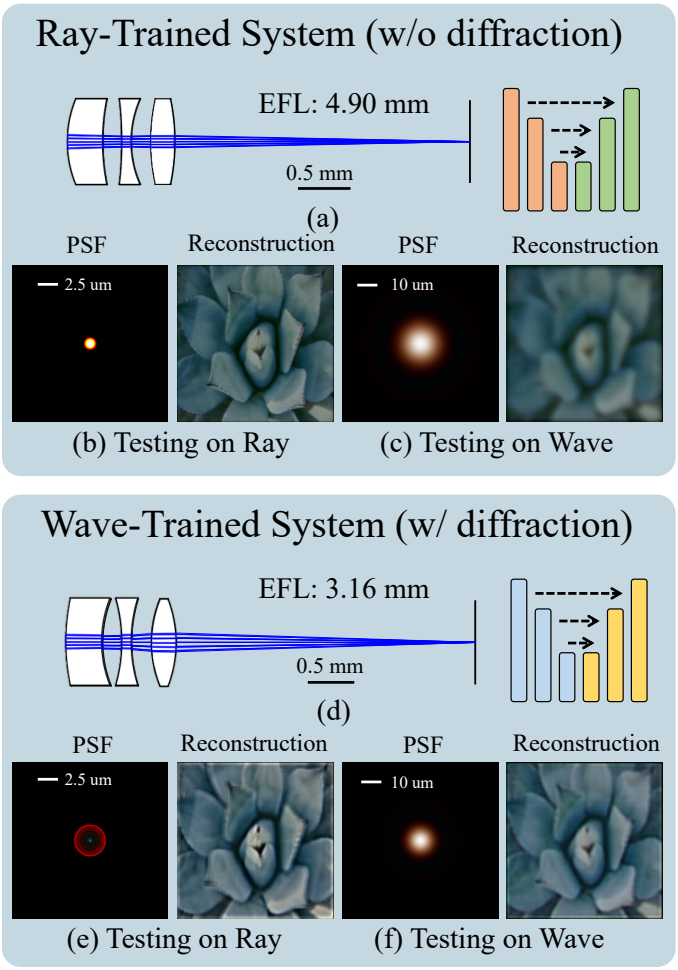


Figure 12. **An analysis of ray-trained and wave-trained Cooke Triplet lens.** With different physics models in end-to-end optimization, Cooke Triplet lens adapts to different architectures. The ray-trained one yields minimal geometric RMS spot size, while the wave-trained one pursues shorter effective focal length for better diffraction performance. When diffraction is introduced to testing, the wave-trained system exhibits better robustness to wave optics effects.

2.3. Cooke Triplet Lens Optimization

Similar to Fig. 7 of the main article, we compare Cooke Triplet lenses optimized using ray optics and wave optics. As observed, the two optimization approaches lead to distinct lens configurations and network adaptations. The ray-

Table 3. Classification accuracy on wave optics rendered measurements

Pixel size (um)	Training physics		MF	CL
	Wave	Ray		
Aperture radius: 0.1 mm				
1	**72.0%**	48.5%	11.67	0.800
5	**72.8%**	65.8%	3.177	0.979
10	**73.6%**	67.7%	0.435	0.999
Aperture radius: 0.3 mm				
1	**71.2%**	63.1%	0.252	0.999
5	**73.4%**	67.5%	0.104	0.999
10	**76.4%**	68.7%	0.120	0.999

optimized system focuses on minimizing the RMS spot size, but because diffraction effects are not accounted for during training, the network remains sensitive to diffraction blur. In contrast, the wave-optimized system mitigates diffraction blur by modifying both the optical design and network architecture, resulting in a shorter effective focal length (EFL). This adaptation not only suppresses diffraction in the PSF but also enables the network to reconstruct intensity from diffraction-blurred measurements. When evaluated under wave-optics-based rendering, the wave-trained system demonstrates greater robustness compared to the ray-trained one.

2.4. Optimized Lens Recipe

We provide the recipe of wave-optimized six layer aspherical lens displayed in Fig. 1 in the main paper in Table 4 and 5 for reference. We include the radius, curvatures, materials, thickness, and aspherical parameters at each surface.

3. Image Classification

To examine the impact of the physics model on semantic information quality in rendering, we conduct end-to-end optimization for image classification. The training dataset consists of 10 image classes from ImageNet [3], with 800 images used for training and 100 images for testing per class. We utilize ResNet18 [4] for class label estimation and employ cross-entropy loss to drive the joint optimization of both the lens and the network.

Specifically, the experiments are conducted on a Cooke Triplet lens across different physics models, aperture radii, and pixel sizes, with classification performance consistently evaluated using measurements rendered with wave optics.

3.1. Quantitative Analysis

As summarized in Table 3, for an aperture radius of 0.1 mm, systems optimized by ray optics and wave optics have noticeable differences in F-number and lens architecture. The ray-trained system experiences a significant drop in classification accuracy when diffraction is accounted for in inference. This performance drop diminishes as pixel size increases, reducing the structural disparity between wave-PSFs and ray-PSFs. Additionally, with a larger aperture radius of 0.3 mm, the differences in F-number and classification performance between ray-trained and wave-trained systems also become less pronounced, reducing the impact of ignoring diffraction during training.

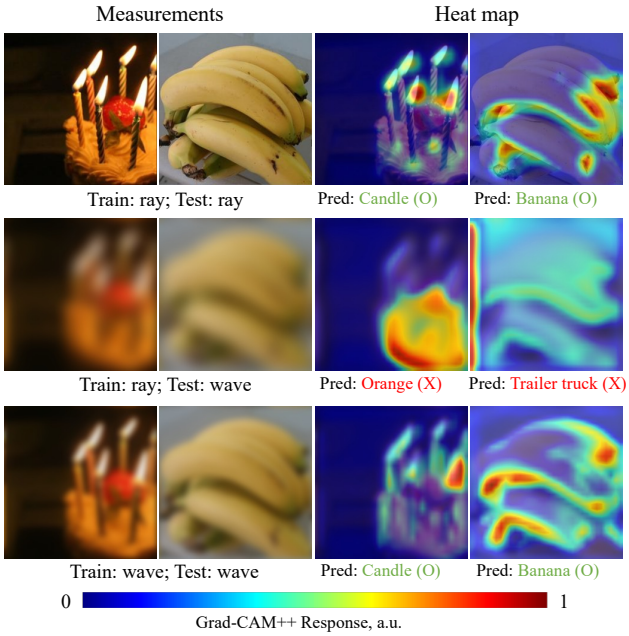


Figure 13. **Examples of classification results and associated heat maps.** For each image, we display its measurement from the system, and associated heat map visualized by GradCAM++. The ray-trained system can only correctly identify key information from sharp measurements generated by ray optics. When we consider diffraction blur, the measurement becomes blurry and the network fails to capture the correct features. On the other hand, the wave-trained system can identify features from measurements suffered from diffraction blur and provide accurate classification.

3.2. Qualitative Analysis

We also conduct a qualitative analysis of the impact of different physics models used for training lenses with a 0.1 mm aperture radius. Specifically, we employ Grad-CAM++ [2], a technique for network interpretation, to visualize feature extraction during classification. In Fig. 13, we present sample measurements alongside their associated heatmaps. As illustrated in the top row, for the ray-trained system, the network is trained to capture features from sharp images. Consequently, when tested with ray optics, the network successfully identifies all necessary features for classification. However, as shown in the middle row, once diffraction blur is introduced to the measurement, the ray-trained network misidentifies features and produces erroneous classification

Index	Type	Radius (mm)	RoC (mm)	k	Material	Thickness (mm)
1	Aperture	0.1	-	-	Air	0.1009
2	Aspheric	1.19	17.6816	4.3395	PMMA	0.8639
3	Aspheric	1.45	-22.5285	8.1054	Air	0.3493
4	Aspheric	1.565	-18.4637	4.5904	OKP4	0.7976
5	Aspheric	1.85	-36.7507	-10.4875	Air	0.1721
6	Aspheric	2.095	26.5022	0.6336	COC	0.8421
7	Aspheric	2.095	-23.6907	-1.6301	Air	0.6343
8	Aspheric	2.065	-23.3454	0.1592	Polycarb	0.4525
9	Aspheric	2.335	-63.9289	-82.4477	Air	0.4747
10	Aspheric	2.35	-79.6274	-19.3968	COC	0.7123
11	Aspheric	2.455	-18.4009	8.2904	Air	0.7348
12	Aspheric	2.585	-14.3779	-0.938	COC	0.9845
13	Aspheric	3.48	13.3459	7.9706	Air	0.9717

Table 4. Lens surface specifications.

Index	a_2	a_4	a_6	a_8	a_{10}	a_{12}
2	-0.0091	-0.0193	-0.0016	-0.0005	-0.0003	-0.0003
3	-0.0416	-0.0187	0.0008	0.0002	0.00002	-0.00002
4	-0.1321	0.0006	0.0012	0.00001	-0.00002	-0.0000006
5	-0.0735	-0.0074	-0.0016	-0.0002	-0.00001	-0.000003
6	0.1032	-0.001	0.0002	0.00005	0.000005	0.000003
7	-0.0709	-0.00008	0.0004	0.00004	0.000003	-0.000002
8	-0.0785	-0.0156	-0.0008	-0.00007	-0.000005	0.00000095
9	-0.0259	0.0006	-0.0009	-0.0001	-0.000007	-0.0000008
10	0.0554	-0.0072	0.00001	0.00001	-0.000015	-0.000004
11	-0.095	0.001	-0.0004	-0.00008	-0.000002	0.000004
12	-0.0079	-0.006	-0.0006	-0.00002	0.000003	0.0000008
13	0.0366	-0.0082	-0.0001	0.00001	0.000002	-0.0000003

Table 5. Aspheric surface coefficients on different lens surfaces.

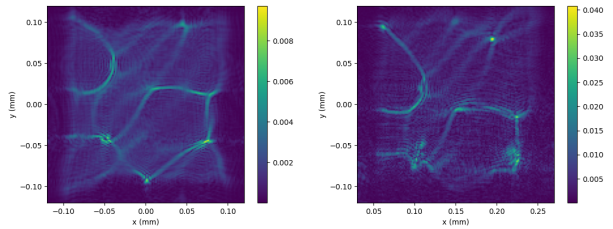
results. In contrast, as shown in the bottom row, when diffraction is considered during optimization, the lenses are driven to a lower F-number to mitigate the diffraction blur, and the network is aware of diffraction in feature extraction. This results in less blurry measurements and more accurate feature detection, as demonstrated in the right column. This analysis reaffirms the critical importance of incorporating diffraction in end-to-end optimization.

4. Free-Form Optics

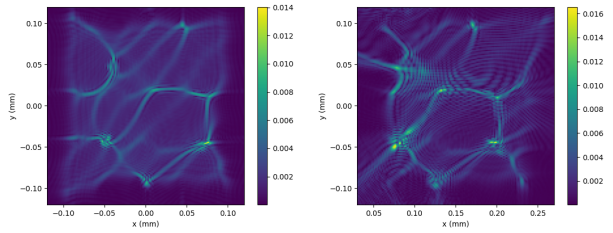
In addition to lens design, we apply our wave optics model to an inverse rendering experiment on freeform optics [1]. Specifically, we illuminate the free form optics surface with coherent plane wave, modeling the interactions between coherent rays, and compute measurements.

4.1. Measurement Rendering

Provided the discrete height map at the given coordinates, we use two geometric profiles, linear meshing, and a B-spline model, to characterize the surface and render PSFs. For each model, we display on-axis and off-axis measurements generated by our simulator in Fig. 14. As we can see, if we directly mesh the surface by linearly interpolating the heights over the surface, visible artifacts appear, especially for off-axis measurements. On the other hand, as observed in Fig. 14c and 14d, by adopting B-Spline as fitting geometry profile to the height maps, we obtain smoother measurements with fewer artifacts. The results demonstrate that the proposed simulator is capable of modeling the structures and off-axis distortions in wave optics propagation in freeform optics.

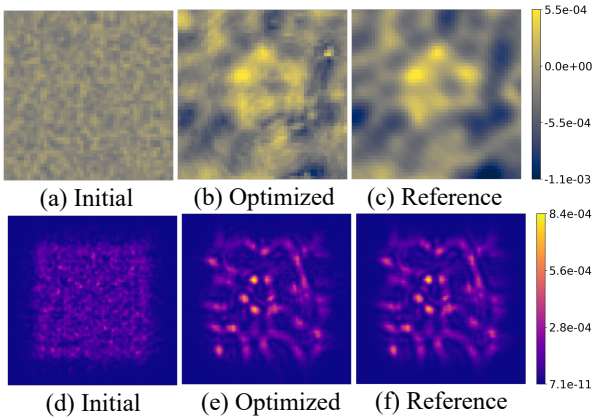


(a) Meshed, On-axis PSF. (b) Meshed, Off-axis PSF.



(c) BSpline, On-axis PSF. (d) BSpline, Off-axis PSF.

Figure 14. **Measurements taken from a freeform optics surface using different geometric profiles in surface fitting.** We illuminate a freeform optical surface and compare non-grid interpolation parameterized by meshing or BSpline. With B-spline parametrization, the freeform optics provides smoother measurements with fewer artifacts.



(a) Initial (b) Optimized (c) Reference

(d) Initial (e) Optimized (f) Reference

Figure 15. **Recovering a freeform optics surface from its measurement.** Setup: A monochromatic plane wave is modulated by a reference freeform optical surface (c), generating the corresponding measurement (f). Starting from a randomly initialized surface (a) and its associated measurement (d), our differentiable wave optics model accurately captures coherent light interference during propagation, enabling precise reconstruction of the surface (b) and measurement (e). Top row: Surface profiles. Bottom row: Corresponding measurements. Measurement field size: $0.3\ \text{mm}^2$. Surface size: $0.24\ \text{mm}^2$.

4.2. Height Map Recovery

We recover the reference height map of a free form optical element through differentiable rendering. In particular, we first take a reference height map, illustrated in Fig. 15 (c), and render its associated measurement visualized in Fig. 15 (f). Next, we start optimizing a randomly initialized optics surface, shown in Fig. 15 (a), with associated measurements shown in Fig. 15 (d). The surface optimization is driven by the mean squared error between rendered and reference measurements.

Figures 15 (b) and 15 (e) present the optimized height maps and the corresponding measurement. As observed, the optimized height map closely aligns with the reference, and the rendered measurements accurately replicate the reference values.

References

[1] Nick Antipa, Grace Kuo, Reinhard Heckel, Ben Mildenhall, Emrah Bostan, Ren Ng, and Laura Waller. Diffusercam: lensless single-exposure 3d imaging. *Optica*, 5(1):1–9, 2017. 7

[2] Aditya Chattopadhay, Anirban Sarkar, Prantik Howlader, and Vineeth N Balasubramanian. Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. In *2018 IEEE winter conference on applications of computer vision (WACV)*, pages 839–847. IEEE, 2018. 6

[3] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 6

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 6