Supplementary Information for: Generalized Aberrations for Processing-Aware Optical Design

Geoffroi Côté[1], Ethan Tseng[1], and Felix Heide[1]

[1]Princeton University, USA

This document offers additional methodological details, insights, and in-depth analyses in support of the findings from the main manuscript. Design files for all the lenses presented in this work are available. Given the comprehensive scope of our modeling and optimization method, we present the accompanying notes as fully self-contained. Sections with key novel contributions or analyses include:

- Section S1.3.2: Differentiable vignetting model that incorporates vignetting factors into the optimization problem, in contrast to established software that treats them as constants.
- Section S1.4.2: Insight into kernel choices for kernel density estimation (KDE) in geometric point spread functions (PSFs) estimation.
- Section S1.4.3: Diffraction heuristic to account for diffraction effects in geometric PSFs.
- Section S2.2.4: Constraints for optimizing glass materials, including modeling the glass catalog with a mesh.
- Section S3.3: Detailed description of our end-to-end (E2E) optimization procedure for E2E image restoration.
- Section S6: Design of a C-mount lens using three strategies: E2E optimization for post-restoration quality, task-driven optimization for pre-restoration quality, and spot size optimization.

Contents

List of Figures

List of Tables

S1 Modeling of Imaging Optics

In this section, we describe the modeling procedure for hybrid optical systems composed of refractive and diffractive components. In the context of task-driven optical design, the goal of the modeling procedure is to accurately simulate the imaging process of a virtual scene—represented by an input image—through the optical system, generating an intermediate image with realistic aberrations. The process must be fully differentiable to enable gradient-based optimization of the lens parameters. Here, we assume that the optical properties of each lens component are known. The parameterization of the lens, which relates the optical properties of each component to the design variables, is discussed in Section S2.1.

At the core of our modeling procedure is optical ray tracing. Rays are characterized by their wavelength λ as well as their position vector $\mathbf{r} = (x, y, z)$ and direction vector $\mathbf{d} = (\alpha, \beta, \gamma)$, with the direction vector normalized to unit length so that $|\mathbf{d}| = 1$. We adopt a sequential ray tracing approach, assuming the order of interaction between rays and optical surfaces is known, and we consider axial symmetry to reduce most operations to functions of the radial coordinate $r = \sqrt{x^2 + y^2}$. Exact ray tracing through the optical system is achieved by alternating between two operations: 1) updating the ray position vectors $\mathbf{r}$ from one optical interface to the next, and 2) updating the ray direction vectors $\mathbf{d}$, typically following Snell's Law. Our primary goal is to obtain the transverse ray coordinates upon hitting the image plane, from which key optical quantities, such as the root-mean-square (RMS) spot size or the geometric PSF, can be computed.

The structure of this section is as follows. In Section S1.1, we discuss dispersion models used to evaluate refractive indices for refractive elements. In Section S1.2, we describe the procedure for propagating and bending rays through individual optical components. In Section S1.3, we extend the procedure to the entire optical system, involving system specifications like aperture and field to obtain spot diagrams. In Section S1.4, we detail our method to evaluate diffraction-compensated geometric PSFs. In Section S1.5, we use geometric PSFs to simulate optical aberrations on input images.

Table S1 provides a summary of the notation used throughout this chapter. Table S2 summarizes the various user-defined parameters that are involved in the process.

S1.1 Modeling of Dispersion

Dispersion is the dependence of the refractive index of a material on the wavelength of light: $n = n(\lambda)$. Dispersion models are used to accurately evaluate the refractive index of a material—either fictitious or real—at different wavelengths, as required to compute Snell's Law.

In modeling glass variables, it is common to consider three quantities: the refractive index at the Fraunhofer line "d" (587.6 nm) n_d, the Abbe number

$$v_\mathrm{d} = \frac{n_\mathrm{d} - 1}{n_\mathrm{F} - n_\mathrm{C}}, \tag{S1}$$

Table S1: Notation used throughout this chapter.

Symbol	Description
Lens characterization	
$n(\lambda)$	Refractive index of a medium
$\mu = {}^{n}/_{n'}$	Refractive index ratio between two adjacent media
n_d	Refractive index at "d" Fraunhofer line (587.6 nm)
v_d	Abbe number defined at "d" Fraunhofer line
$\Delta P_\mathrm{g,F}$	Deviation from normal partial dispersion
$\phi(r)$	Phase profile of a diffractive surface
$\tilde{z}(r)$	Profile of an optical interface
Ray tracing through components	
$\mathbf{r} = (x, y, z)$	Position vector of a ray
$\mathbf{d} = (\alpha, \beta, \gamma)$	Direction vector of a ray
λ	Wavelength of a ray
$r = \sqrt{x^2 + y^2}$	Radial coordinate of a ray
θ	Angle of incidence at an interface
θ'	Angle of refraction at an interface
t	Ray-marching distance
Ray tracing through system	
u_h, where $h = \{1, \ldots, n_\mathrm{h}\}$	Relative field angles
λ_w, where $w = \{1, \ldots, n_\mathrm{w}\}$	Wavelengths
(x_p, y_p), where $p = \{1, \ldots, n_\mathrm{p}\}$	Relative pupil coordinates
$n_\mathrm{r} = n_\mathrm{h} n_\mathrm{w} n_\mathrm{p}$	Number of rays
EPD	Entrance pupil diameter
FOV	Field of view
HFOV	Half field of view
$V_\mathrm{i}^{(h,w)}$, $V_\mathrm{s}^{(h,w)}$, $V_\mathrm{l}^{(h,w)}$	Inferior, superior, and lateral vignetting factors
SR_h	Root-mean-square spot radius at field h
Geometric PSFs	
$W_\mathrm{p} \times H_\mathrm{p}$	Bin count of sampled geometric PSFs
$s_\mathrm{x}, s_\mathrm{y}$	Size of sampled PSFs in absolute units
$\mathbf{K} \in \mathbb{R}^{W_\mathrm{p} \times H_\mathrm{p} \times n_\mathrm{f} \times n_\mathrm{w}}$	Matrix of sampled geometric PSFs
Image simulation	
$W_\mathrm{g} \times H_\mathrm{g}$	Layout of image patches and corresponding PSFs
$W_\mathrm{r} \times H_\mathrm{r}$	Size of rescaled PSFs in the PSF grid
$\mathbf{K}_\mathrm{grid} \in \mathbb{R}^{W_\mathrm{g} \times H_\mathrm{g} \times 3 \times H_\mathrm{r} \times H_\mathrm{r}}$	Matrix representing the PSF grid
$\mathbf{I}$	Input RGB image
$\mathbf{I}'$	Output RGB image

Table S2: Summary of user-defined parameters for image simulation, excluding the lens parameterization covered in Section S2.1. In general, most parameters should be defined as to smoothly capture the variation of optical response with the aperture, field, and wavelength.

Parameter	Description
Lens modeling	
$n(\lambda, n_\mathrm{d}, v_\mathrm{d}, \Delta P_\mathrm{g,F})$	Dispersion model to relate material parameters to refractive indices at wavelength λ
Ray tracing	
EPD	Entrance pupil diameter, can be retrieved from f-number
$(x_1, y_1), (x_2, y_2), \ldots, (x_{n_\mathrm{p}}, y_{n_\mathrm{p}})$	Relative transverse coordinates at the pupil (between -1 and 1) Also referred to as the pupil sampling scheme (see n_rings) Relative coordinates are scaled to the EPD
HFOV	Half field of view
$u_1, u_2, \ldots, u_{n_\mathrm{h}}$	Relative field angles (between 0 and 1), often uniformly spaced Relative field angles are scaled to the half FOV
$\lambda_1, \lambda_2, \ldots, \lambda_{n_\mathrm{w}}$	Wavelengths
PSF	
$W_\mathrm{p} \times H_\mathrm{p}$	Bin count of the sampled geometric PSFs, should be large enough to capture the vast majority of traced rays
$s_\mathrm{x}, s_\mathrm{y}$	Size in absolute units of the sampled geometric PSFs, typically set to match the pixel pitch of the sensor
σ	Support of the kernel for KDE (e.g., cosine kernel), typically twice the bin size
Image simulation	
$W_\mathrm{g} \times H_\mathrm{g}$	Layout of image patches and PSFs in the PSF grid
$\mathbf{W}_\mathrm{w}$	Weight matrix for combining PSFs at different wavelengths
OP	Overlap percentage between image patches

and the partial dispersion

$$P_{\mathrm{g,F}} = \frac{n_{\mathrm{g}} - n_{\mathrm{F}}}{n_{\mathrm{F}} - n_{\mathrm{C}}}\,, \tag{S2}$$

where n_{C}, n_{d}, n_{F}, and n_{g} are the refractive indices at the Fraunhofer lines "C" (656.3 nm), "d" (587.6 nm), "F" (486.1 nm), and "g" (435.8 nm), respectively.

For most refractive materials, $P_{\mathrm{g,F}}$ is approximately linear with v_{d} [27] so that we can express it as a function of the deviation from so-called "normal partial dispersion" $\Delta P_{\mathrm{g,F}}$,

$$P_{\mathrm{g,F}} = 0.6438 - 0.001682 v_{\mathrm{d}} + \Delta P_{\mathrm{g,F}}\,. \tag{S3}$$

As a result, when optimizing the glass parameters, we can choose to optimize only n_{d} and v_{d}—in which case we assume that only materials of normal partial dispersion $\Delta P_{\mathrm{g,F}} = 0$ are available—or, optionally, $\Delta P_{\mathrm{g,F}}$ as well.

In this work, we use the Hartmann dispersion model [17],

$$n(\lambda) = A + \frac{C}{\lambda - B}\,, \tag{S4}$$

where A, B, and C can be evaluated from the definitions of n_{d}, v_{d}, and $P_{\mathrm{g,F}}$, with the latter given by Equation (S3), as follows:

$$B = \frac{-\lambda_{\mathrm{C}}\lambda_{\mathrm{F}} + \lambda_{\mathrm{C}}\lambda_{\mathrm{g}} + P_{\mathrm{g,F}}\left(\lambda_{\mathrm{C}}\lambda_{\mathrm{g}} - \lambda_{\mathrm{F}}\lambda_{\mathrm{g}}\right)}{-\lambda_{\mathrm{F}} + \lambda_{\mathrm{g}} + P_{\mathrm{g,F}}\left(\lambda_{\mathrm{C}} - \lambda_{\mathrm{F}}\right)}\,; \tag{S5}$$

$$C = \frac{\left(B - \lambda_{\mathrm{C}}\right)\left(B - \lambda_{\mathrm{F}}\right)\left(n_{\mathrm{d}} - 1\right)}{v_{\mathrm{d}}\left(\lambda_{\mathrm{C}} - \lambda_{\mathrm{F}}\right)}\,; \tag{S6}$$

$$A = \frac{B n_{\mathrm{d}} + C - \lambda_{\mathrm{d}} n_{\mathrm{d}}}{B - \lambda_{\mathrm{d}}}\,. \tag{S7}$$

S1.2 Ray Tracing Through Optical Components

In this section, we describe the procedure for tracing rays through individual optical components, focusing on the update equations for the position vectors $\mathbf{r}$ and direction vectors $\mathbf{d}$.

S1.2.1 Propagation and Ray-Marching Distance

The propagation of a ray refers to the operation that updates the ray position vector $\mathbf{r}$ from one optical interface to the next. The propagation operation is defined by

$$\mathbf{r}' = \mathbf{r} + t\mathbf{d}\,, \tag{S8}$$

where t, the ray-marching distance, is the only unknown. Under axial symmetry, the profile of the interface is given by a function $\tilde{z}(r)$ that returns the sag $\tilde{z}$ of the interface at a given radial position r. Finding the ray-marching distance is generally an iterative process that involves solving

the following equation [31]:

$$f(\mathbf{r}) = \tilde{z}(r) - z = 0\,. \tag{S9}$$

In practice, when closed-form solutions are not available, the ray-marching distance can be evaluated iteratively using Newton's method,

$$t^{(k+1)} = t^{(k)} - \frac{f(\mathbf{r} + t^{(k)}\mathbf{d})}{\partial f(\mathbf{r} + t^{(k)}\mathbf{d})/\partial t^{(k)}}\,, \tag{S10}$$

where $t^{(k)}$ denotes the ray-marching distance estimate at iteration k. The derivative term can be computed from the analytical expression of $\tilde{z}(r)$. Different parameterizations for the interface profile $\tilde{z}(r)$ are discussed in Section S2.1.3. Optical interfaces often use simple analytical expressions for $\tilde{z}(r)$ such as polynomial expressions, which facilitate the analytical evaluation of the derivative term. For spherical or flat interfaces, the ray-marching distance can be determined using a closed-form expression, eliminating the need for an iterative procedure.

Failure cases A failure case arises with *backtracking* rays, where $t < 0$ but $\gamma > 0$—the ray direction is forward, but the ray-marching distance is negative. This suggests the next interface is behind the current one, indicating overlapping optical interfaces. In this scenario, the ray remains well-behaved even though the solution is not physically meaningful, and we continue tracing it rather than discarding it. To prevent this case during optimization, a constraint can be added to ensure that $t > 0$, as discussed in Section S2.2.2. Another failure case occurs when no solution exists for Equation (S9), indicating that the ray does not intersect the interface, leading to the ray being discarded. To mitigate missed surfaces during optimization, soft constraints on the angle of incidence can be employed, as explained in Section S2.2.2. A third failure case happens in iterative ray tracing only, when Newton's method fails to converge or returns a ray-marching distance corresponding to a radial coordinate r where $\tilde{z}(r)$ is undefined. In this case, the ray is also discarded.

S1.2.2 Refractive Interfaces

Snell's Law describes the refraction of a ray at refractive interfaces,

$$\sin(\theta') = \mu \sin(\theta)\,, \tag{S11}$$

where $\mu = {}^{n}/_{n'}$ is the refractive index ratio of the media on either side of the interface, and θ and θ' are the angles between the rays and the normal to the interface. To explicitly recover the update equation for direction vectors $\mathbf{d}$, Snell's Law can be written in vector form:

$$\mathbf{d}' = \mu\mathbf{d} + \big(\cos(\theta') - \mu\cos(\theta)\big)\,\mathbf{n}\,, \tag{S12}$$

where $\mathbf{n}$ is the normal to the interface. Recognizing that $\cos^2(\theta) = \gamma^2$, the square of the z-component of $\mathbf{d}$, we can rewrite Equation (S11) as

$$\cos^2(\theta') = 1 - \mu^2 \left(1 - \gamma^2\right) . \tag{S13}$$

The normal to the surface $\mathbf{n}$ is proportional to the gradient of the implicit surface function $f(\mathbf{r})$ from Equation (S9),

$$\mathbf{n} \propto -\nabla f = -\frac{\partial \tilde{z}}{\partial x}\mathbf{i} - \frac{\partial \tilde{z}}{\partial y}\mathbf{j} + \mathbf{k} , \tag{S14}$$

and is normalized so that $|\mathbf{n}| = 1$. In general, the expression for aspheric surfaces is given in a simple and standard analytical form, and it is common to derive an analytical expression for the gradient term as in Section S1.2.1.

Failures cases When the expression for Equation (S13) is negative, the cosine of the angle of refraction is imaginary, indicating total internal reflection. In this case, the ray is discarded. To prevent total internal reflection, a constraint can be added to the optimization problem to ensure a positive result for Equation (S13), as described in Section S2.2.2.

S1.2.3 Diffractive Interfaces

Assuming axial symmetry, tracing rays through diffractive interfaces and surfaces modeled by a phase profile—such as metasurfaces—can be achieved using the generalized law of refraction [37],

$$n' \sin(\theta') - n \sin(\theta) = \frac{\lambda}{2\pi} \frac{\partial \phi}{\partial r} , \tag{S15}$$

where ϕ is the phase profile of the interface at radial coordinate r and λ is the wavelength.

The derivative term in Equation (S15) is known as the *phase gradient*, which can be computed analytically or numerically. The preferred method is to estimate the phase gradient pointwise using automatic differentiation, requiring $\phi(r)$ to be parameterized continuously. To accurately represent the behavior of real diffractive surfaces, the phase profile $\phi(r)$ must be appropriately parameterized or constrained, as further discussed in Section S2.1.4.

Assuming air on both sides of the surface ($n' = n = 1$), the updated direction cosines can be retrieved as follows [7]:

$$\alpha' = \alpha + \frac{\lambda}{2\pi} \frac{\partial \phi}{\partial r} \frac{x}{r} = \frac{\lambda x}{\pi} \frac{\partial \phi}{\partial r^2} ; \tag{S16}$$

$$\beta' = \beta + \frac{\lambda}{2\pi} \frac{\partial \phi}{\partial r} \frac{y}{r} = \frac{\lambda y}{\pi} \frac{\partial \phi}{\partial r^2} ; \tag{S17}$$

$$\gamma' = \sqrt{1 - \alpha'^2 - \beta'^2} , \tag{S18}$$

where the phase gradient is computed in a numerically stable manner to avoid division by zero.

It is important to note that the generalized law of refraction does not perfectly describe the behavior of rays through diffractive interfaces and generally agrees with diffraction theory only

under certain conditions [14, 24, 26, 29], particularly when phase discontinuities at the boundaries of diffraction zones are multiples of 2π.

Failures cases Similar to refractive interfaces, a failure case arises when the argument of the square root term in Equation (S18) is negative, resulting in an undefined updated direction vector. For instance, this occurs at a flat interface under normal incidence when $\partial\phi/\partial r = 2\pi/\lambda$. To prevent this scenario during the optimization process, a constraint can be added to the optimization objective to ensure that the argument of the square root term remains non-negative.

S1.3 Ray Tracing Through the Optical System

This section details the procedure for tracing rays through the entire optical system, involving system specifications, with the field and aperture specifications being the most critical. In conventional camera systems, the field specification is typically defined in terms of the field of view (FOV), while the aperture specification is described by the f-number, which is the ratio of the effective focal length (EFL) to the entrance pupil diameter (EPD).

S1.3.1 Ray Initialization

For on-axis imaging, incoming ray bundles from object space must completely fill the aperture stop of the lens. The objective of the ray initialization process is to achieve this condition while adequately handling off-axis ray bundles. The step-by-step procedure for initializing rays is as follows:

1. Locate the entrance pupil using a paraxial ray-tracing operation.
2. Specify the EPD either directly or derive it from the f-number, requiring the EFL to be known or calculated from a paraxial operation.
3. Initialize the ray position vectors $\mathbf{r}$ to uniformly fill the entrance pupil at n_p different relative intersections (x_p, y_p), as shown in Figure S1, referred to as the *pupil sampling scheme*. The axial coordinate $z = z_\mathrm{p}$ corresponds to the entrance pupil's location relative to the first optical surface of the lens.
4. Replicate the same rays in n_h different relative fields u_h, and initialize the ray direction vectors $\mathbf{d}$ with $\alpha = 0$, $\beta = \sin(u_h \cdot \mathrm{HFOV})$, and $\gamma = \cos(u_h \cdot \mathrm{HFOV})$. A common approach is to uniformly generate u_h between 0 and 1. This document assumes that fields are always expressed in terms of field angles.
5. Replicate the ray position and direction vectors for all user-specified wavelengths $\lambda_1, \lambda_2, \ldots, \lambda_{n_\mathrm{w}}$.
6. If vignetting is present, estimate the vignetting factors. To account for pupil aberrations when the aperture stop is not in object space, further refine the vignetting factors if necessary.

Update the lateral coordinates of the ray position vectors $\mathbf{r}$ according to the vignetting factors as described in Section S1.3.2.

In practice, ray-tracing operations are batched over $n_\mathrm{r} = n_\mathrm{h} n_\mathrm{w} n_\mathrm{p}$ rays, where n_h, n_w, and n_p denote the number of field values, wavelengths, and pupil coordinates, respectively.

S1.3.2 Vignetting and Ray Aiming

For on-axis light, ray bundles should fill the entire clear area of the aperture stop. Since the entrance pupil is a conjugate plane of the aperture stop, rays that uniformly fill the entrance pupil will also adequately fill the aperture stop under the paraxial approximation.

In contrast, off-axis light introduces complexity due to pupil aberrations. Rays initialized at the entrance pupil may deviate significantly from their corresponding positions on the aperture stop, leading to skewed ray-tracing results. Additionally, rays can be obstructed by surfaces other than the aperture stop, such as mechanical stops, causing vignetting. To account for both pupil aberrations and vignetting, it is common to initialize ray bundles at the entrance pupil and adjust them into an elliptical shape that departs from the original entrance pupil configuration.

Under axial symmetry, we define three vignetting factors for each field h and wavelength w: the inferior and superior vignetting factors $V_\mathrm{i}^{(h,w)}$ and $V_\mathrm{s}^{(h,w)}$, and the lateral vignetting factor $V_\mathrm{l}^{(h,w)}$. The inferior and superior vignetting factors represent the fraction of the bottom and top halves of the entrance pupil that is blocked, respectively, while the lateral factor accounts for the fraction of the left and right halves, which are equal due to symmetry. For example, if $V_\mathrm{i} = 1$ and $V_\mathrm{s} = 0$, light will pass freely through the center and top of the entrance pupil but will be blocked between the bottom and center. By definition, all on-axis light passes through the aperture stop unimpeded, thus $V_\mathrm{i} = V_\mathrm{s} = V_\mathrm{l} = 0$.

The relative coordinates x_pupil and y_pupil at the entrance pupil (normalized between -1 and 1) are updated as follows:

$$x'_\mathrm{pupil} = \left(1 - V_\mathrm{l}^{(h,w)}\right) x_\mathrm{pupil}\,; \tag{S19}$$

$$y'_\mathrm{pupil} = \left(1 - \frac{V_\mathrm{i}^{(h,w)} + V_\mathrm{s}^{(h,w)}}{2}\right) y_\mathrm{pupil} + \frac{V_\mathrm{s}^{(h,w)} - V_\mathrm{i}^{(h,w)}}{2}\,. \tag{S20}$$

Vignetting Vignetting occurs when rays from an off-axis field are physically blocked by surfaces other than the aperture stop. These surfaces may include mechanical stops or other optical components with insufficient diameter to allow the rays to pass through. While vignetting reduces relative illumination, it can also enhance optical performance by aiding in the correction of off-axis aberrations.

Although vignetting is influenced by the design's geometry, optical design software often treats vignetting factors as fixed specifications, similar to FOV or aperture, remaining unchanged during optimization. This approach has a significant limitation: the vignetting factors may not accurately represent the actual vignetting of the lens. In particular, even with mechanical stops, achieving a

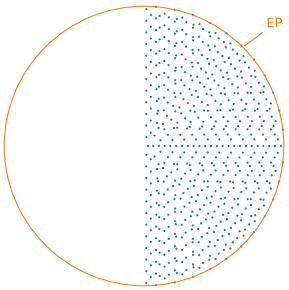


Figure S1: Default pupil sampling scheme. Rays are initialized on concentric rings index by $i = \{1, 2, \ldots, n_{\text{rings}}\}$, with each ring containing $2i - 1$ rays. This results in a total of $n_{\text{r}} = n_{\text{rings}}^2$ rays on one half of the entrance pupil (EP). To prevent aliasing effects and ensure accurate sampling of the outer edge of the pupil, jittering is applied to the ray positions. Due to axial symmetry, only one half of the aperture needs to be traced. In this illustration, $n_{\text{rings}} = 24$.

lens arrangement that faithfully reproduces the specified vignetting factors is not always possible. As a result, recalculating the vignetting factors before each optimization cycle is common practice.

In Section S4.1.5,we discuss a specific implementation of vignetting for telephoto lens design. We consider adding a front aperture in object space, assuming the incoming light is only clipped at the bottom of the front aperture, leading to $V_i > 0$ and $V_s = 0$. This assumption is justified because the entrance pupil is typically located inside the lens and, by convention, off-axis light travels upward. We compute $V_i^{(h,w)}$ for each field based on basic geometry and approximate the lateral vignetting $V_l^{(h,w)}$ to ensure that rays do not pass outside the superior portion of the entrance pupil. This implementation is physically plausible, as light is clipped by the aperture stop, integrating the vignetting factors into the optimization problem. In contrast, conventional commercial software typically treats these factors as fixed parameters. Note that placing mechanical stops outside of object space would necessitate a more complex implementation.

Ray aiming Pupil aberrations can be addressed through an accurate ray-aiming procedure that refines the vignetting factors, ensuring that rays at the entrance pupil effectively span the aperture stop for any given field or wavelength.

In the absence of vignetting, the initial vignetting factors are set to $V_{\text{i}}^{(h,w)} = V_{\text{s}}^{(h,w)} = V_{\text{l}}^{(h,w)} = 0$ for all fields h and wavelengths w. Following the approach outlined in [4], our goal is to refine these factors so that marginal meridional or sagittal rays accurately intersect the edge of the aperture stop. We approximate the required displacements of the inferior $\Delta y_{\text{pupil,i}}^{(h,w)}$, superior $\Delta y_{\text{pupil,s}}^{(h,w)}$, and lateral $\Delta x_{\text{pupil,l}}^{(h,w)}$ rays at the entrance pupil.

During the iterative refinement process, we assume a linear relationship between the entrance

pupil coordinates x_{pupil}, y_{pupil} and the aperture stop coordinates x_{stop}, y_{stop}:

$$\Delta x_{\text{stop}} \approx \Delta x_{\text{pupil}} \frac{\mathrm{d}x_{\text{stop}}}{\mathrm{d}x_{\text{pupil}}}\,; \tag{S21}$$

$$\Delta y_{\text{stop}} \approx \Delta y_{\text{pupil}} \frac{\mathrm{d}y_{\text{stop}}}{\mathrm{d}y_{\text{pupil}}}\,. \tag{S22}$$

We evaluate the ray-aiming errors $\Delta x^{(h,w)}_{\text{stop,l}}$, $\Delta y^{(h,w)}_{\text{stop,i}}$, and $\Delta y^{(h,w)}_{\text{stop,s}}$ by tracing a sagittal ray and two meridional rays (inferior and superior) for each field h and wavelength w, then comparing their intersections with the aperture stop diameter, which is obtained by tracing an on-axis meridional ray. The derivatives required for this calculation can be determined using automatic differentiation. Subsequently, Equation (S21) and Equation (S22) are utilized to compute the necessary displacements at the entrance pupil. For most lenses with moderate pupil aberrations, a single ray-aiming step suffices to refine the ray bundle shape. If greater accuracy is needed, multiple iterations can be performed.

Finally, when accounting for vignetting, we initially establish the vignetting factors and subsequently refine them using the ray-aiming procedure. This approach may introduce minor errors in the vignetting factors, as it does not guarantee that rays are precisely clipped at the edges of the mechanical stops. However, it is generally adequate for most optical systems.

S1.3.3 Spot Diagrams

Spot diagrams are essential for evaluating a lens's optical performance and play a crucial role in computing both the effective spot radius (ESR) and geometric PSF, the latter discussed in Section S1.4.

Following the procedure outlined in Section S1.3.1, n_{r} rays are initialized at the entrance pupil and traced through the optical system to the image plane. There, we obtain spot diagrams, which consist of n_{r} lateral coordinates $x_{h,w,p}$ and $y_{h,w,p}$, where h, w, and p are the field, wavelength, and pupil indices, respectively. From the spot diagrams, we can compute the field-wise RMS spot radius

$$\text{SR}_h = \sqrt{\frac{1}{n_{\text{p}} \sum_w W_w} \sum_{w,p} W_w \left(\left(x_{h,w,p} - \overline{x}_h\right)^2 + \left(y_{h,w,p} - \overline{y}_h\right)^2 \right)}\,, \tag{S23}$$

where $\overline{x}_h$ and $\overline{y}_h$ are the centroids of the spot diagrams for field h, with $\overline{x}_h = 0$ due to axial symmetry and

$$\overline{y_h} = \frac{1}{n_{\text{p}}} \frac{\sum_{w,p} W_w y_{h,w,p}}{\sum_w W_w}\,, \tag{S24}$$

with W_w representing user-defined weights for each wavelength. Rays experiencing ray failure are excluded from the calculations of the centroid $\overline{y_h}$ and spot radius SR_h.

S1.3.4 Validation of Ray Tracing

We validate our ray-tracing procedure by comparing it with the commercial optical design software Synopsys CODE V. The lens under consideration is a smartphone telephoto lens, as illustrated in Figure S2, with a 14.0 mm EFL, a $\pm 9.5^\circ$ FOV, and an f-number of 2.8. This all-refractive lens incorporates four aspheric elements, making it well-suited for validating the ray-tracing process for refractive interfaces. To ensure consistency in dispersion models, we compute the refractive indices for each material at the relevant wavelengths within our framework before exporting the lens design to CODE V. This allows for a direct comparison of results between our method and CODE V.

In Figure S3, we present the ray fan plots generated by both our framework and CODE V, revealing a close correspondence between the two sets of results. Additionally, Table S3 compares the spot size diameter—defined as twice the RMS spot size radius calculated using Equation (S23)—with values obtained from CODE V. This comparison is based on nine equally spaced wavelengths between 450–650 nm with uniform weights. The results indicate that the spot size diameter aligns within ± 0.001 µm across all fields and wavelengths.

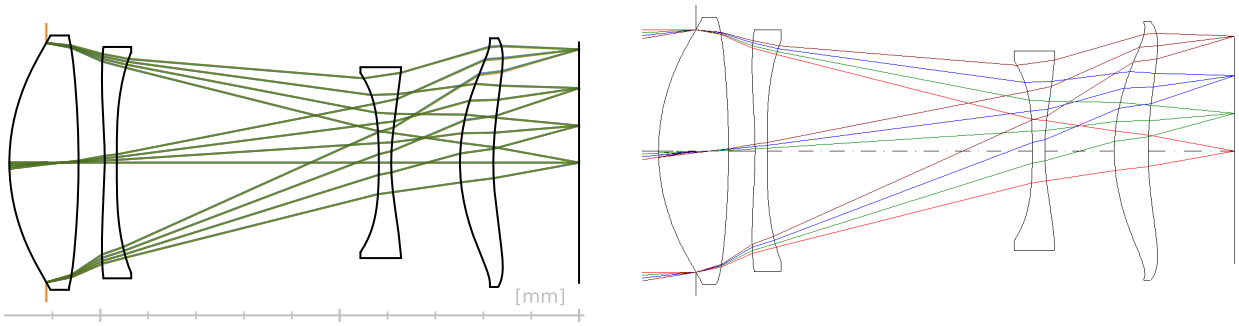
Table S3: Spot size diameter for the 4-piece telephoto smartphone lens, obtained from our framework and CODE V for the same fields and wavelengths. The spot size diameter is averaged from 9 uniformly spread wavelengths between 450–650 nm with uniform weights. The agreement between the two methods is on the order of ± 0.001 µm. Note that this comparison employs higher ray sampling ($n_{\text{rings}} = 256$) than what we typically use for image simulation. In practice, we note that an aperture stop located within the lens rather than in object space, as well as the inclusion of mechanical stops, may lead to larger discrepancies in the spot size diameter than those observed here.

Field (°)	0	0.95	1.9	2.85	3.8	4.75	5.7	6.65	7.6	8.55	9.5
Ours (µm)	1.479	1.485	1.500	1.519	1.534	1.538	1.540	1.557	1.609	1.685	1.716
CODE V (µm)	1.480	1.485	1.500	1.519	1.534	1.538	1.540	1.557	1.609	1.685	1.716

S1.4 Point Spread Function Estimation

The PSF characterizes the distribution of light intensity in the image plane when a point source is imaged, making it a crucial parameter for simulating the imaging process. This section outlines the procedure for estimating the PSFs of a lens system based on the spot diagrams obtained in Section S1.3.3.

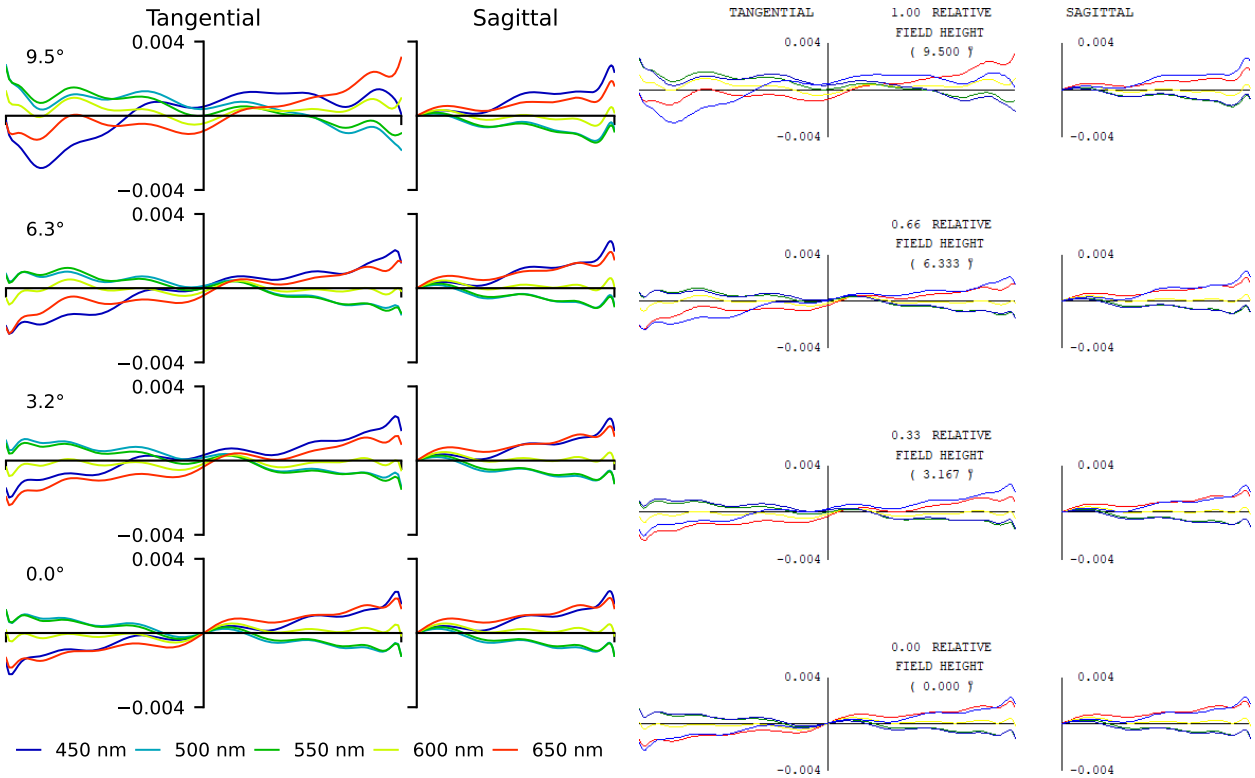
In Section S1.4.1, we detail the ray-counting method, which serves as the foundation for geometric PSF estimation. In Section S1.4.2, we introduce the KDE method, a smooth and differentiable alternative to ray counting that allows gradient-based optimization. In Section S1.4.3, we propose a heuristic solution to partially account for diffraction effects.



(a) 2D Layout from Our Framework

(b) 2D Layout from CODE V

Figure S2: Layout plots for a 4-piece telephoto smartphone lens with a 14.0 mm EFL, a $\pm 9.5^\circ$ FOV, and an f-number of 2.8, obtained from (a) our framework and (b) CODE V. There is no vignetting in this particular lens, and ray aiming is unnecessary as the aperture stop is located in object space.



(a) Ray Fan Plots, Our Framework

(b) Ray Fan Plots, CODE V

Figure S3: Ray fan plots for the 4-piece telephoto smartphone lens, obtained from (a) our framework and (b) CODE V for the same fields and wavelengths. To obtain tangential ray fan plots, we initialize rays across a vertical line from the bottom of the entrance pupil to the top, and plot their tangential ray aberration at the image plane. For sagittal ray fan plots, the rays are initialized across a horizontal line from the center of the entrance pupil to the rightmost point, and the sagittal ray aberration is illustrated. The two sets of ray fan plots are visually indistinguishable.

S1.4.1 Ray-Counting Method

Under dominant geometric aberrations, diffraction can be safely ignored, and the PSFs can be computed using the ray-counting method: setting a virtual detector at the image plane and counting the rays that hit each bin [6], as obtained from the spot diagrams.

The outcome of the PSF estimation is represented by a matrix $\mathbf{K} \in \mathbb{R}^{n_\mathrm{f} \times W_\mathrm{p} \times H_\mathrm{p} \times n_\mathrm{w}}$ where n_f and n_w are the number of fields and wavelengths, and W_p and H_p are the width and height of the virtual grid, respectively. This grid, initialized for each field h, is centered on the centroid of the spot diagram $(\overline{x}_h, \overline{y}_h)$, which corresponds to the centroid of the deposited energy.

We typically set the physical size of each bin to match that of a pixel in the image sensor, eliminating the need for interpolation. The size of the grid $W_\mathrm{p} \times H_\mathrm{p}$ is selected such that the total physical area $(s_\mathrm{x}, s_\mathrm{y})$ is sufficient to capture the majority of rays during optimization across all fields and wavelengths. The virtual grid is populated by counting the number of rays that hit each bin. The resulting PSFs are then normalized by the total number of traced rays, ensuring that the total energy is conserved; that is, the sum of the PSF values for a given field and wavelength equals one.

It is important to note that this method does not account for rays that completely miss the virtual sensor. If not addressed, this omission could artificially inflate the perceived imaging properties as, in reality, rays that miss the sensor would contribute to a loss of contrast due to scattering. To mitigate this, we redistribute the energy of such rays evenly across all bins. This approach effectively simulates scattering, discouraging the optimization process from favoring rays that miss the virtual sensor.

S1.4.2 Kernel Density Estimation

Naive ray counting is non-differentiable relative to the transverse ray coordinates, violating the requirement for gradient-based optimization. To address this, we employ the differentiable alternative of KDE, which effectively spreads the energy of each ray across multiple bins. We utilize a 2D kernel with a finite support size σ, which corresponds to the range where the kernel has a non-zero value.

Among kernels with finite support, we focus on separable kernels, which can be expressed as the product of two 1D kernels. Additionally, we are interested in kernels that exhibit antisymmetry about $\pm\sigma/4$ within the range $[0, \pm\sigma/2]$. These properties ensure that each ray distributes an equal amount of energy across the bins, independent of where it intersects the grid, provided that the support size is an even multiple of the bin size. This characteristic simplifies the estimation of energy not accounted for by rays that miss the virtual sensor, thereby eliminating the need for normalization of the PSF.

Two examples of 2D kernels with these desirable properties are the triangular and cosine kernels, as visualized in Figure S4. In our experiments, we employ the triangular kernel; however, we note that the cosine kernel produces similar results.

A natural choice for σ is to set it to twice the size of a bin. This configuration ensures that a

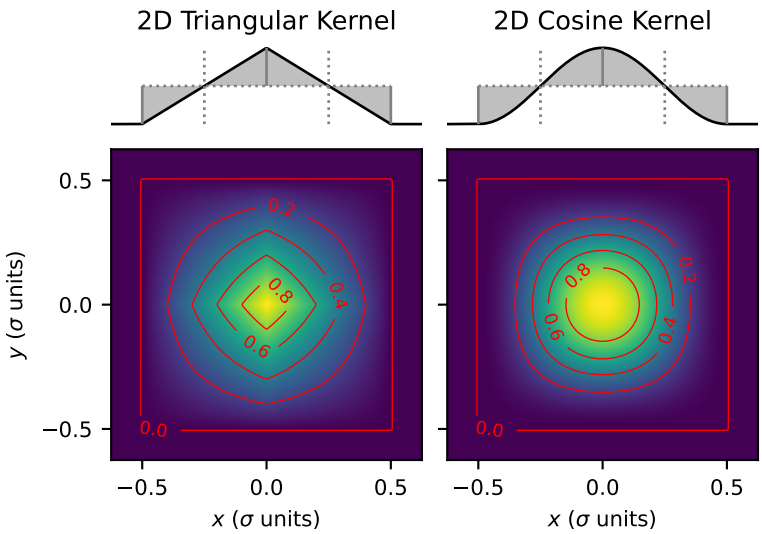


Figure S4: The triangular and cosine kernels are effective candidates for KDE, as they spread the energy of each ray across multiple bins. Both kernels possess a finite support size, and due to the antisymmetry of their corresponding 1D kernels around $\pm\sigma/4$, the resulting 2D kernels ensure that each ray distributes an equal amount of energy across the bins. This uniform distribution occurs regardless of the specific location where a ray intersects the grid, provided the support size is an even multiple of the bin size. This eliminates the need for PSF normalization, thereby simplifying the estimation of energy not accounted for by rays that miss the virtual sensor.

ray's energy is distributed across four bins in two dimensions, except in the discrete case where a ray precisely intersects the center of a bin. Choosing a larger support size can be beneficial when the PSFs are expected to be significantly larger than a single bin, which reduces the need for higher ray sampling rates. Conversely, selecting a smaller support size is inadvisable, as it would result in some rays concentrating their energy within a single bin, compromising the differentiability property.

S1.4.3 Accounting for Diffraction Effects

Geometric PSFs do not account for diffraction effects, which can be significant in many optical systems. To optimize image quality in lens systems, it is crucial to prevent the optimization process from favoring sub-diffraction-limited PSFs that are not physically realizable. Such solutions can lead to biased results and waste degrees of freedom in the optimization problem. To address this, we introduce a heuristic to partially account for diffraction effects.

Motivation In establishing our diffraction heuristic, we first make the following observations. When the geometric PSF is well within the diffraction-limited regime, the diffraction PSF is generally *larger* than the geometric counterpart. Such is the case when all rays hit the same coordinates at the image sensor, forming a Dirac delta geometric PSF. In this scenario, the intensity PSF matches the Airy disk's intensity distribution for on-axis light, and is usually larger or similar for off-axis light due to a decrease in effective f-number and relative illumination. In contrast, when the geometric

PSF size is close to or slightly larger than the Airy disk, the diffraction PSF tends to be *narrower* due to destructive interference, which benefits image quality. Finally, for geometric PSFs significantly larger than the Airy disk, the diffraction PSF size often aligns with the geometric counterpart.

Diffraction heuristic design Based on these observations, our proposed diffraction heuristic aims to account for the finite size of the Airy pattern while considering destructive interference effects to prevent underestimation or overestimation of PSF size. The principle of our heuristic approximates that each ray in the spot diagram generates its own Airy field pattern, with equal energy distributed across the pattern, offset by the ray's lateral coordinates at the image plane. These field patterns interfere, allowing for both constructive and destructive interference.

Airy disk field pattern generation We generate an Airy disk field pattern for each wavelength. The Airy disk pattern for a circular aperture, defined by the numerical aperture (numerical aperture (NA)) related to the f-number by $\mathrm{NA} = 1/2f\#$, is expressed as follows [38]:

$$\mathrm{U}_{\mathrm{Airy}}(r) \propto \frac{2J_1(k\mathrm{NA}r)}{k\mathrm{NA}r}\,, \tag{S25}$$

where J_1 is the Bessel function of the first kind (order 1), $k = 2\pi/\lambda$ is the wavenumber, and $r = \sqrt{x^2 + y^2}$ is the radial distance. The radius of the first dark ring of the Airy disk is approximately $r_1 \approx 0.61\lambda/\mathrm{NA}$. For the discrete kernel corresponding to the Airy field pattern, we set its size $S_{\mathrm{Airy}} \geq 6r_1$ for all wavelengths of interest. The discrete diffraction patterns are represented as $\mathbf{K}_{\mathrm{diffraction}} \in \mathbb{R}^{S_{\mathrm{Airy}} \times S_{\mathrm{Airy}} \times n_{\mathrm{w}}}$.

Diffraction-compensated PSF estimation To reconstruct the field pattern for the geometric PSF, we empirically find that assuming equal phase everywhere yields satisfactory results. We extract the square root of the geometric PSFs, convolve the approximated field with the Airy field pattern, and square the result to recover the intensity. In the case of a Dirac delta geometric PSF, the final intensity PSF matches the Airy disk intensity pattern, as expected. While this procedure does not replace a full diffraction model, it effectively accounts for diffraction effects. Most importantly, it ensures that the optimization process does not favor sub-diffraction-limited PSFs that are not physically realizable.

S1.5 Imaging Simulation

In this section, we outline the procedure for simulating physically realistic geometric aberrations on an input RGB image I, which serves to represent a virtual scene.

The differentiable lens simulation model applies a spatially varying convolution to the input RGB image I, yielding the aberrated image

$$I'(x', y') \approx \mathrm{PSF}(x', y') \otimes I(x', y')\,, \tag{S26}$$

where $\otimes$ denotes the convolution operation. We model Equation (S26) by discretizing the portion

of the sensor corresponding to the image capture into a grid of $W_g \times H_g$ regions. Each region is associated with its respective PSF, and the convolution is performed by applying the corresponding PSF to the corresponding region of the input image. The collection of $W_g \times H_g$ PSFs is referred to as the *PSF grid* as described in Section S1.5.1.

S1.5.1 PSF Grid

To perform the spatially varying convolution, we first need to construct the PSF grid, where each PSF element corresponds to the optical response for a specific portion of the sensor.

The complete procedure to obtain the PSF grid is outlined below:

1. From the spot diagrams, the diffraction-compensated PSFs are evaluated at n_h equidistant field values h and for n_w wavelengths as described in Section S1.4: $\mathbf{K}_1 \in \mathbb{R}^{n_f \times W_p \times H_p \times n_w}$.

2. In the general case, the PSFs are composed of $n_w \neq 3$ wavelengths that must be combined to represent three color channels (RGB). A used-defined weight matrix $\mathbf{W}_w \in \mathbb{R}^{n_w \times 3}$ should be generated based on the spectral sensitivity of the RGB channels, with each column summing to 1. The PSFs at different wavelengths are combined according to a matrix multiplication on the last dimension of both tensors: $\mathbf{K}_2 = \mathbf{K}_1 \mathbf{W}_w$, with $\mathbf{K}_2 \in \mathbb{R}^{n_f \times W_p \times H_p \times 3}$. An example of such a weight matrix is provided in Section S4.1.2 for the telephoto lens study.

3. Likewise, the PSFs initially represent n_h fields that must be recombined into $W_g \times H_g$ sensor regions. A weight matrix $\mathbf{W}_f \in \mathbb{R}^{n_f \times (W_g H_g)}$ is created to specify how the field-wise PSFs should be weighted to generate the PSF grid elements. As with the previous step, each column must sum to 1. Each PSF in the grid is a weighted average of the sampled PSFs, where the weight for a field h corresponds to the proportion of the sensor region that is closest to it. We obtain the intermediate result $\mathbf{K}_3 \in \mathbb{R}^{W_g \times H_g \times W_p \times H_p \times 3}$ by applying the matrix multiplication $\mathbf{K}_3 = \mathbf{K}_2 \mathbf{W}_f$ on the first dimension of $\mathbf{K}$. In contrast to naive interpolation, this weighted average scheme involves the full FOV of the lens in the simulation and optimization pipeline.

4. The interpolated PSFs are then rotated to the appropriate angle using three consecutive shear operations [35], preserving the energy of the PSFs in contrast to naive interpolation. We obtain $\mathbf{K}_4 = \text{rotate}(\mathbf{K}_3)$, with $\mathbf{K}_4 \in \mathbb{R}^{W_g \times H_g \times W_p \times H_p \times 3}$. Note that a portion of the energy can be lost in the rotation due to the finite support of the PSFs. In such cases, similar to what was discussed in Section S1.4, we redistribute the lost energy equally across all bins of the PSF element.

5. (Optional) If the physical size of pixels on the virtual sensor and image does not match, the PSF grid is resized to size $W_r \times H_r$ to match the image resolution. The result $\mathbf{K}_5 \in \mathbb{R}^{W_g \times H_g \times W_r \times H_r \times 3}$ is obtained from $\mathbf{K}_4$ by bilinear interpolation. In this case, the PSFs are normalized again so that the area for a given grid item and color channel is 1.

The PSF grid $\mathbf{K}_{\text{grid}} \in \mathbb{R}^{W_g \times H_g \times W_r \times H_r \times 3}$ is obtained as a result of this series of operations.

We note that this procedure is also applicable to image patches, i.e., the case where the simulated image corresponds only to a region of the sensor instead of its entirety.

S1.5.2 Spatially Varying Convolution

We employ the spatially varying overlap-add method [11]. The image is split into $W_g \times H_g$ patches that overlap on one another so that the borders represent a proportion of the original patch size called the overlap percentage (OP), e.g., 25 %. Each patch is convolved with its corresponding PSF, taken from the PSF grid $\mathbf{K}_{\text{grid}}$. The resulting convolved patches are reconstructed to form the final image. When there is overlap, the reconstructed image is a pixel-wise weighted average of the convolved patches. For smooth interpolation, we use a 2D Hann window for the weighting scheme, so that the regions between image patches progressively interpolate between adjacent PSFs.

S2 Optimization of Imaging Optics

This section outlines the process of optical design optimization within our method, focusing on three key components: parameterization, optimization objective, and optimization algorithm. Since different design tasks require tailored solutions, this section does not prescribe a universal approach but instead presents options that can be combined and customized for specific problems. We begin with definitions.

The set of all design variables is denoted by $\boldsymbol{\theta} \in \mathbb{R}^N$, where each individual variable is represented as θ_n. These variables include optimizable elements such as spacings, glass variables, curvatures, aspheric coefficients, and parameters for diffractive surfaces. The process of defining the set of lens variables for optimization, along with their relationship to the overall lens characterization, is referred to as *parameterization* and is detailed in Section S2.1.

To establish the optimization objective, we introduce a vector-valued function $\boldsymbol{\ell}(\boldsymbol{\theta}) \in \mathbb{R}^M$, which maps the design variables to a set of *residuals* $\ell_m(\boldsymbol{\theta})$. These residuals serve distinct purposes: some measure optical performance or image quality criteria to be minimized, while others enforce constraints that must be satisfied. The residuals and their roles are discussed further in Section S2.2.

In the absence of hard constraints, the optimization objective is typically expressed as a sum of squares

$$L(\boldsymbol{\theta}) = \frac{1}{2} \|\boldsymbol{\ell}\|^2 , \tag{S27}$$

representing a least-squares problem when $M \geq N$. The goal of unconstrained optimization is to find the set of variables $\boldsymbol{\theta}^*$ that minimizes L:

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} L(\boldsymbol{\theta}) . \tag{S28}$$

The methods used to solve this optimization problem are described in Section S2.3.

S2.1 Parameterization

Lens parameterization defines how the lens variables $\boldsymbol{\theta} \in \mathbb{R}^N$ are structured and how they correspond to the lens characterization. In this section, we discuss the parameterization of each category of lens variables, as summarized in Table S4. In addition, a summary of parameterization options is provided in Table S5.

S2.1.1 Spacings

Spacings are the distances between adjacent surfaces such as the aperture stop, refractive or diffractive interfaces, or the image plane. We parameterize spacings s in a dimensionless form

$$s' = \alpha^{-1} s , \tag{S29}$$

Table S4: Summary of the parameterization of lens variables, including each category of variables and their scaled counterparts. The scale factor α is a user-defined parameter in units of distance.

Variable type	Variable	Scaled version
Spacings	s'	$s = \alpha s'$
Glass variables	$\mathbf{g}$	$[n_\mathrm{d} \vert v_\mathrm{d}] = \mathbf{W}_\mathrm{g}^{-1} \left(\mathbf{g} - \mathbf{b}_\mathrm{g}\right)$
Curvatures	c'	$c = {}^1/_\alpha c'$
Aspheric parameter vector even-asphere model; κ is conic constant	$\mathbf{a}' = [\kappa, a'_4, a'_6, \ldots]$	$\mathbf{a} = [\kappa, \alpha^{-4} a'_4, \alpha^{-6} a'_6, \ldots]$
Diffractive parameter vector	$\mathbf{d}' = [d'_2, d'_4, \ldots]$	$\mathbf{d} = [\alpha^{-2} d'_2, \alpha^{-4} d'_4, \ldots]$

Table S5: Summary of parameterization options and optimization features that can facilitate the optimization process depending on the problem.

Feature	Description
Last airspace solve	Solve for the last airspace to enforce a given property; only one type of solve can be active
TTL solve	Enforce a given total track length (TTL); the last airspace is removed from the optimization problem
Curvature solve	Solve for a user-specified curvature to enforce a given property; only one type of solve can be active; the curvature is removed from the optimization problem
EFL solve	Enforce a given effective focal length (EFL)
Image height solve	Enforce a given paraxial image height; the surface used for the solve should be past the aperture stop
Bézier parameterization	Apply a linear transformation on the aspheric coefficients to partially disentangle them, similar to control points on a Bézier curve

where α is a user-defined scale factor in units of distance, typically set globally for all variables. A common choice for α is half the entrance pupil diameter (EPD), representative of the half-diameter of most lens surfaces. The scale factor impacts gradient magnitudes and optimization step sizes, though nearly scale-invariant methods like the Levenberg-Marquardt (LM) algorithm (Section S2.3.2) mitigate this effect. Equation (S29) allows conversion between lens variables and their scaled parameters—a feature common to all parameterizations in this section.

Total track length (TTL) solve Optionally, a solve can be implemented to enforce a specified TTL by solving for the last airspace. This removes the last airspace from the optimization problem. This option is used in the telephoto lens case study (Section S4), where fixing the TTL with a hard constraint is desirable.

S2.1.2 Refractive Materials

To model a glass's dispersion curve, we use the refractive index n_d at the "d" Fraunhofer line (587.6 nm), the Abbe number v_d, and optionally, the deviation from partial dispersion $\Delta P_{\mathrm{g,F}}$ (see Section S1.1). These two or three parameters fully describe the dispersion curve. Instead of using their raw forms, we normalize them as a vector $\mathbf{g} \in \mathbb{R}^{d_\mathrm{glass}}$, where d_glass represents the number of glass parameters (2 or 3).

The practical aim of glass variable optimization is to select the optimal materials from the set of *catalog glasses* $\mathcal{G} = \{g'_1, g'_2, \ldots, g'_{n_\mathrm{cat}}\}$ available for the lens design problem. Since glass materials exist in discrete form, a continuous relaxation is often employed to explore the solution space.

Following the approach of Côté et al. [6], we define an intermediate representation for glass variables using a linear transformation fitted to a whitening transformation of the catalog glasses:

$$\mathbf{g} = \mathbf{W}_\mathrm{g}[n_\mathrm{d}|v_\mathrm{d}] + \mathbf{b}_\mathrm{g}\,, \tag{S30}$$

where $\mathbf{W}_\mathrm{g} \in \mathbb{R}^{d_\mathrm{glass} \times d_\mathrm{glass}}$ is the whitening matrix, $\mathbf{b}_\mathrm{g} \in \mathbb{R}^{d_\mathrm{glass}}$ is the offset, and $[\cdot|\cdot]$ denotes concatenation, with $\mathbf{g} \in \mathbb{R}^{d_\mathrm{glass}}$. Note that $\Delta P_{\mathrm{g,F}}$ is included in Equation (S30) when three glass parameters are used. In practice, $\mathbf{W}_\mathrm{g}$ and $\mathbf{b}_\mathrm{g}$ are fitted using the materials available for the optimization, with catalog glasses in sparsely populated regions receiving higher weight, and those in dense regions receiving less weight.

S2.1.3 Refractive Surfaces

The profile of an axially symmetric refractive surface is fully characterized by its height $\tilde{z}(r)$ as a function of the radial position r [28]:

$$\tilde{z}(r) = \frac{cr^2}{1 + \sqrt{1 - (1+\kappa)(cr)^2}} + f(r)\,, \tag{S31}$$

where the first term accounts for the curvature c and conic constant κ, while $f(r)$ describes the aspheric contribution, often as a polynomial. Special cases include: $f(r) = 0$ for a conic surface, $f(r) = \kappa = 0$ for a spherical surface, and $f(r) = c = 0$ for a flat surface. The profile is defined for all r only if $\kappa \leq -1$, otherwise restricted to $r^2 \leq 1/(1+\kappa)c^2$.

Spherical surface parameterization For spherical surfaces, the profile depends solely on the curvature c, the inverse of the radius. A dimensionless curvature is defined as:

$$c' = \alpha c \,, \tag{S32}$$

where α is a user-defined scale factor in units of distance.

Aspheric surface parameterization Aspheric surfaces are defined by the curvature c and additional parameters, grouped into a vector $\mathbf{a}$, which includes the conic constant κ and aspheric coefficients a_i. The aspheric contribution $f(r)$ is commonly modeled using the even-asphere formulation, defined as

$$f(r) = \sum_{i=2}^{n+1} a_{2i} r^{2i} \,, \tag{S33}$$

where a_i are the aspheric coefficients and n is the user-defined number of coefficients. No term in r^2 is included to avoid correlation with curvature. Likewise, r^4 terms correlate with κ, which is why κ is often fixed and frozen to 0 or -1. Setting $\kappa = -1$ improves numerical stability, ensuring the square root argument in Equation (S31) remains positive.

The normalized parameter vector, used for parameterization, is labeled $\mathbf{a}'$ and consists of the dimensionless coefficients

$$a'_{2i} = \alpha^{2i} a_{2i} \,. \tag{S34}$$

We note that the conic constant κ is dimensionless, so we do not apply a scaling factor to it.

Alternative asphere models Other asphere models, such as Q-type polynomials [10], can improve optimization by reducing parameter correlations. We empirically find that applying a linear transformation on the aspheric coefficients can improve convergence speed by disentangling the coefficients. As it is a linear transformation, it can be converted back-and-forth into the original even-asphere model. These transformations are applied only during parameterization and reverted to the even-asphere model Equation (S33) prior to ray tracing. In experiments, we apply a linear transformation similar to a Bézier curve parameterization: rather than optimizing the coefficients directly, this option allows a quantity similar to control points on the surface to be optimized.

Focal length solve To enforce a desired effective focal length (EFL), the curvature c can be solved via paraxial ray tracing. Any refractive surface can be selected as long as the refractive index ratio between the two media differs from unity.

Paraxial image height solve For applications like those in Section S4, maintaining a fixed paraxial image height is preferable to maintaining a fixed focal length. When defocus is zero, a fixed EFL

ensures consistent image height. However, non-zero defocus causes variations in image height, problematic for systems with fixed sensor sizes and field of view (FOV). To address this, we define the image-height-consistency focal length, which accounts for defocus to maintain a fixed paraxial image height. This value matches the EFL at zero defocus and can be enforced by solving for the curvature of any optical surface beyond the aperture stop.

S2.1.4 Diffractive Surfaces

Diffractive surfaces are typically parameterized by their phase profile, where it is assumed that all light is diffracted into the desired order. Once defined, the phase profile is used along with the generalized law of refraction to calculate the output angle of incident rays, as detailed in Section S1.2.3. The phase profile is parameterized using the common polynomial form

$$\phi(r) = \sum_{i=1}^{n} d_{2i} r^{2i}\,, \tag{S35}$$

where d_{2i} denote each of the n phase profile parameters. Note that the phase profile is wavelength-independent, which results in the typical negative dispersion of diffractive surfaces, with an equivalent Abbe number of -3.45.

The dimensionless diffractive parameter vector $\mathbf{d}'$ is related to the original vector $\mathbf{d}$ as:

$$d'_{2i} = \alpha^{2i} d_{2i}\,. \tag{S36}$$

S2.2 Optimization Objective

The unconstrained optimization objective $L(\boldsymbol{\theta}) = 1/2 \left\|\boldsymbol{\ell}\right\|^2$ is the sum of squares of the individual residuals ℓ_m. In optical design, the optimization objective generally varies between applications depending on the lens configuration and constraints that must be satisfied. This section provides an overview of different types of residuals employed in our method to construct optimization objectives.

In Table S6, we summarize the residuals used in the framework, including their number of elements and user-defined parameters. The presence and weight of constraints should be set based on the specific problem. These residuals represent a subset of possible soft constraints; additional constraints can be added as needed.

S2.2.1 Optical Performance

In conventional design, as opposed to task-driven design, we consider the transverse ray aberrations (TRA) optical performance metric, which closely relates to the root-mean-square (RMS) spot size.

To compute TRA, we first generate spot diagrams as described in Section S1.3.3. Rays are initialized at n_h fields, n_w wavelengths, and n_p positions at the entrance pupil, and traced from the object space to the image plane, where their (x, y) transverse coordinates are obtained.

Table S6: Summary of the categories of residuals, including their maximum number of terms and user-defined parameters. The maximum number of terms may scale with the number of rays n_r, spacings n_s, interfaces n_i, refractive elements n_g, or fields n_h.

Residuals	Description	Maximum number of terms	User-defined parameters
ℓ_TRA	Transverse ray aberration residuals (Transverse errors at image plane)	$2n_\mathrm{r}$	–
ℓ_RP	Ray path residuals (Boundaries on path along optical axis)	$n_\mathrm{s}n_\mathrm{r}$	$t^{(k)}_\mathrm{z,min}$, $t^{(k)}_\mathrm{z,max}$ for all k
ℓ_RA	Ray angle residuals (Maximum incidence/refraction angles)	$2n_\mathrm{i}n_\mathrm{r}$	Maximum θ_max
ℓ_SN	Surface normal residuals (Maximum angle between ray and normal)	$n_\mathrm{i}n_\mathrm{r}$	Maximum θ_max
ℓ_D	Distortion residuals (Maximum distortion for any field)	n_h	Maximum D_max
ℓ_RI	Relative illumination residuals (Minimum relative illumination for any field)	n_h	Minimum R_min
ℓ_GMD	Glass mesh distance residuals (Distance from glass variables to glass catalog mesh)	n_g	Maximum edge length
ℓ_G	Glass catalog distance residuals (Distance to nearest catalog glass)	n_g	–

The TRA are the differences between these coordinates and the centroids $\overline{x}_h, \overline{y}_h$ for each field h. These centroids are averages over all wavelengths w and positions p,

$$\Delta x_{hwp} = x_{hwp} - \overline{x}_h = x_{hwp}\,; \tag{S37}$$

$$\Delta y_{hwp} = y_{hwp} - \overline{y}_h = y_{hwp} - \frac{1}{n_\mathrm{p}} \frac{\sum_{w,p} W_w y_{h,w,p}}{\sum_w W_w}\,, \tag{S38}$$

where W_w is the user-defined weight for wavelength w. Additionally, ray failures are excluded from centroid calculations.

The residual vector $\boldsymbol{\ell}_\mathrm{TRA}$ corresponding to TRA contains $2n_\mathrm{r}$ elements, where n_r is the number of traced rays, and is defined as

$$\boldsymbol{\ell}_\mathrm{TRA} = \frac{1}{\sqrt{n_\mathrm{r}}}[\Delta x_1, \Delta y_1, \ldots, \Delta x_{n_\mathrm{r}}, \Delta y_{n_\mathrm{r}}]\,, \tag{S39}$$

with combined ray indices shown for simplicity. The scaling term ensures that the sum-of-squares objective $1/2\,\|\boldsymbol{\ell}_\mathrm{TRA}\|^2$ remains independent of the number of rays. Non-uniform weights for wavelengths, field angles, or pupil positions are also possible, as demonstrated over wavelengths in Section S4.1.2 for the telephoto lens study.

Minimizing the sum-of-squares objective involving TRA is equivalent to minimizing the square

of the RMS spot size, provided there is only one field h and no ray failures:

$$\|\boldsymbol{\ell}_{\text{TRA}}\|^2 = \frac{1}{n_{\text{w}} n_{\text{p}}} \sum_{p=1}^{n_{\text{p}}} \sum_{w=1}^{n_{\text{w}}} \left(\Delta x_{wp}^2 + \Delta y_{wp}^2 \right) . \tag{S40}$$

S2.2.2 Geometric Constraints

These constraints employ the concept of using rays as probes as in [6]. Rays traced for spot diagram evaluation provide valuable information about lens geometry, which we use to enforce constraints based on intermediate operands from each ray r, such as ray-marching distances or angles of incidence at interfaces.

We define the ramp function $\text{ramp}(x) = \max(x, 0)$, which ensures that constraints influence the optimization process only when violated, i.e., when $x > 0$.

Ray path residuals The *ray path residuals* ℓ_{RP} aim to avoid overlapping surfaces, enforce sufficient center/edge thicknesses in refractive elements, and impose a sufficient image clearance (the clear space between the last element and the image sensor). These residuals are derived from the ray-marching distance projected onto the optical axis t_z across every glass or air spacing k.

For each optical spacing k, we define user-specified lower and upper bounds $t_{\text{z,min}}^{(k)}$ and $t_{\text{z,max}}^{(k)}$. Setting a lower threshold of at least 0 for all spacings, including airspaces, avoids backtracking rays. In refractive elements, the lower and upper thresholds enforce a minimum and maximum thickness. Thresholds can be also be dynamically defined relative to the center thickness to constain the edge-to-thickness ratios.

The penalty term for a ray r traveling through spacing k is

$$\Delta z = \text{ramp}\left(\max\left(t_{\text{z}} - t_{\text{z,max}}, t_{\text{z,min}} - t_{\text{z}}\right)\right) , \tag{S41}$$

where $\Delta z = 0$ indicates compliance with the bounds, and $\Delta z > 0$ signifies a violation.

With n_{s} spacings in the optical system, the ray path residual vector concatenates up to $n_{\text{s}} \times n_{\text{r}}$ individual residuals,

$$\boldsymbol{\ell}_{\text{RP}} = \frac{1}{\sqrt{n_{\text{r}}}} \left[\Delta z_1, \Delta z_2, \ldots, \Delta z_{n_{\text{s}} \times n_{\text{r}}}\right] , \tag{S42}$$

where the scaling term ensures the residual magnitude is independent of the number of rays. Typically, a well-behaved optical system has $\Delta z = 0$ for most rays, allowing efficient pruning of inactive terms during gradient and Jacobian computations.

Ray angle residuals The *ray angle residuals* ℓ_{RA} constrain the angles of incidence θ and refraction θ' within a user-defined range $\pm\theta_{\text{max}}$, e.g., $\pm 60^\circ$. This aims to stabilize optimization, improve tolerancing, and prevent ray failures, such as missed surfaces or total internal reflection.

Intermediate quantities are defined as $\zeta_{\text{I}} = \cos^2(\theta)$ and $\zeta_{\text{R}} = \cos^2(\theta')$, where negative values indicate missed surfaces for incidence and total internal reflection for refraction. The penalties for

each ray r at interface k are:

$$\Delta\zeta_{\mathrm{I}} = \mathrm{ramp}\left(\cos^2\left(\theta_{\max}\right) - \zeta_{\mathrm{I}}\right) \; ; \tag{S43}$$

$$\Delta\zeta_{\mathrm{R}} = \mathrm{ramp}\left(\cos^2\left(\theta_{\max}\right) - \zeta_{\mathrm{R}}\right) \; . \tag{S44}$$

These penalties are positive if the angle exceeds the allowable range.

With n_{i} interfaces, the ray angle residual vector concatenates up to $2n_{\mathrm{i}} \times n_{\mathrm{r}}$ residuals,

$$\boldsymbol{\ell}_{\mathrm{RA}} = \frac{1}{\sqrt{n_{\mathrm{r}}}} \left[\Delta\zeta_{\mathrm{I},1}, \Delta\zeta_{\mathrm{R},1}, \Delta\zeta_{\mathrm{I},2}, \Delta\zeta_{\mathrm{R},2}, \ldots, \Delta\zeta_{\mathrm{I},n_{\mathrm{i}}\times n_{\mathrm{r}}}, \Delta\zeta_{\mathrm{R},n_{\mathrm{i}}\times n_{\mathrm{r}}}\right] . \tag{S45}$$

Surface normal residuals The *surface normal residuals* ℓ_{SN} encourage manufacturability by restricting the angle between refractive interface normals and the optical axis to $\pm\theta_{\max}$ (e.g., $\pm 30^\circ$). Mirroring the ray angle residuals, we define $\zeta_{\mathrm{SN}} = \cos^2(\theta_{\mathrm{SN}})$, where θ_{SN} is the angle between the surface normal and the optical axis.

The penalty term for ray r at surface k is

$$\Delta\zeta_{\mathrm{SN}} = \mathrm{ramp}\left(\cos^2\left(\theta_{\max}\right) - \zeta_{\mathrm{SN}}\right) \; , \tag{S46}$$

which is non-zero if the surface normal exceeds the allowable angular range.

With n_{i} interfaces, the surface normal residual vector concatenates up to $n_{\mathrm{i}} \times n_{\mathrm{r}}$ individual residuals,

$$\boldsymbol{\ell}_{\mathrm{SN}} = \frac{1}{\sqrt{n_{\mathrm{r}}}} \left[\Delta\zeta_{\mathrm{SN},1}, \Delta\zeta_{\mathrm{SN},2}, \ldots, \Delta\zeta_{\mathrm{SN},n_{\mathrm{i}}\times n_{\mathrm{r}}}\right] . \tag{S47}$$

S2.2.3 Imaging Constraints

These constraints enforce desirable imaging properties beyond optical performance, such as distortion control and relative illumination.

Distortion residuals The *distortion residuals* limit lens distortion to within a user-defined range $\pm D_{\max}$ (e.g., $\pm 2\,\%$). For a given field h, distortion is defined as the relative difference between the centroid $\overline{y}_h$ of the spot diagram and the paraxial image height y'_h:

$$D_h = \frac{\overline{y}_h - y'_h}{y'_{n_{\mathrm{h}}}} \, , \tag{S48}$$

where $y'_{n_{\mathrm{h}}}$ is the paraxial image height of the outermost field. The residuals penalize distortion exceeding $D_{\max}$:

$$\Delta D_h = \mathrm{ramp}\left(|D_h| - D_{\max}\right) \; , \tag{S49}$$

and the distortion residual vector concatenates up to n_{h} individual residuals:

$$\boldsymbol{\ell}_{\mathrm{D}} = \frac{1}{\sqrt{n_{\mathrm{h}}}} \left[\Delta D_1, \Delta D_2, \ldots, \Delta D_{n_{\mathrm{h}}}\right] . \tag{S50}$$

Relative illumination residuals The *relative illumination residuals* ensure the relative illumination across all fields remains above a threshold R_{min} (e.g., 70 %). This constraint is particularly relevant in systems with mechanical stops, as in the telephoto lens study of Section S4, to prevent the optimization process from introducing excessive vignetting.

A common approach to estimate relative illumination, as proposed in Rimmer [22], uses the direction cosines of three specific rays (inferior, superior, and lateral) traced through the system for each field h. The relative illumination is then:

$$R_h = \frac{\alpha_{\text{l},h} \left(\beta_{\text{i},h} - \beta_{\text{s},h}\right)}{2\beta^2_{\text{s},h=1}} , \tag{S51}$$

where $\alpha_{\text{l},h}$, $\beta_{\text{i},h}$, and $\beta_{\text{s},h}$ are the direction cosines of the lateral, inferior, and superior rays, respectively, and $\beta_{\text{s},h=1}$ is the direction cosine of the on-axis meridional ray.

While Equation (S51) provides a good approximation of the relative illumination, it considers only three rays per field, allowing optimization to artificially improve illumination by manipulating just these rays when enough degrees of freedom are available. To address this, we refine the estimation by incorporating all rays in the spot diagram. Defining $\alpha_{\text{std},h}$ and $\beta_{\text{std},h}$ as the standard deviations of the lateral and meridional direction cosines for field h, optionally weighted by wavelength, the relative illumination is updated as:

$$R_h = \frac{\alpha_{\text{std},h}\beta_{\text{std},h}}{\alpha_{\text{std},h=1}\beta_{\text{std},h=1}} . \tag{S52}$$

We empirically find that the estimated relative illumination from Equation (S52) better correlates with established optical design software like Synopsys CODE V.

Residuals penalize fields where relative illumination falls below the threshold:

$$\Delta R_h = \text{ramp}\left(R_{\text{min}} - R_h\right) , \tag{S53}$$

and the relative vector concatenates up to n_{h} individual residuals,

$$\boldsymbol{\ell}_{\text{RI}} = \frac{1}{\sqrt{n_{\text{h}}}} \left[\Delta R_1, \Delta R_2, \ldots, \Delta R_{n_{\text{h}}}\right] . \tag{S54}$$

S2.2.4 Glass Material Constraints

The soft constraints for glass materials ensure that optimized glass variables align with a known catalog of materials. This approach prevents optimization from producing non-physical solutions while enabling sufficient flexibility for exploration.

Glass mesh distance residuals The *glass mesh distance residuals* constrain the glass variables $\boldsymbol{g}$ to remain near the catalog of available materials. If the variables deviate too far from realistic values, optimization may explore impractical solutions, wasting effort and missing optimal designs.

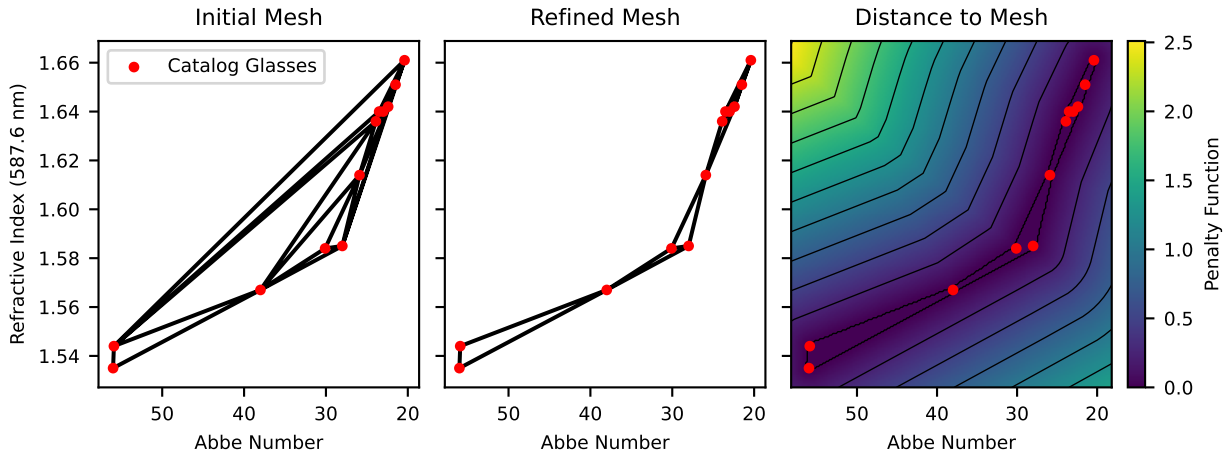


Figure S5: Generation of the glass mesh and calculation of the glass mesh distance residuals, for the 2-dimensional case. The left panel shows the initial mesh generated using Delaunay tessellation. The middle panel shows the refined mesh after removing simplices with excessively large edges. The right panel illustrates the glass mesh distance residuals, where the color indicates the distance to the nearest point on the mesh.

The glass catalog, typically comprising 2- or 3-dimensional variables (e.g., refractive index and Abbe number), is represented as a mesh in the dimensionless glass variable space. This mesh is created using Delaunay tessellation [16], forming triangles in 2D or tetrahedra in 3D. Initially, the mesh's boundary is the convex hull of the catalog, which can be overly coarse and impose loose constraints (left panel of Figure S5 for the moldable materials of the telephoto lens study with 2-dimensional variables).

To refine the mesh, simplices (triangles/tetrahedra) with excessively long edges are removed. A user-defined threshold (e.g., 1.25) specifies the maximum edge length, normalized in glass variable space. If removing all simplices connected to a vertex would isolate it, the simplest simplex (smallest maximum edge length) is retained. The refined mesh offers tighter constraints (middle panel of Figure S5).

The penalty function $L_{\mathrm{GMD}}(\mathbf{g})$ is defined based on the Euclidean distance between a glass variable $\mathbf{g}$ and the nearest point on the mesh. This penalty is zero if $\mathbf{g}$ lies within the mesh (right panel of Figure S5). In 3D cases, a similar process applies, as illustrated in Figure S6 for the microscope objective lens (MOL) study.

The glass mesh distance residuals for all n_{g} glass variables are:

$$\ell_{\mathrm{GMD}} = [L_{\mathrm{GMD}}(\mathbf{g}_1), L_{\mathrm{GMD}}(\mathbf{g}_2), \ldots, L_{\mathrm{GMD}}(\mathbf{g}_{n_{\mathrm{g}}})]\,. \tag{S55}$$

Glass catalog distance residuals While the glass mesh distance residuals prevent non-physical solutions, they do not enforce commitment to specific catalog materials. To address this, the *glass catalog distance residuals* encourage the glass variables to approach the nearest material in the catalog, as illustrated in Figure S7.

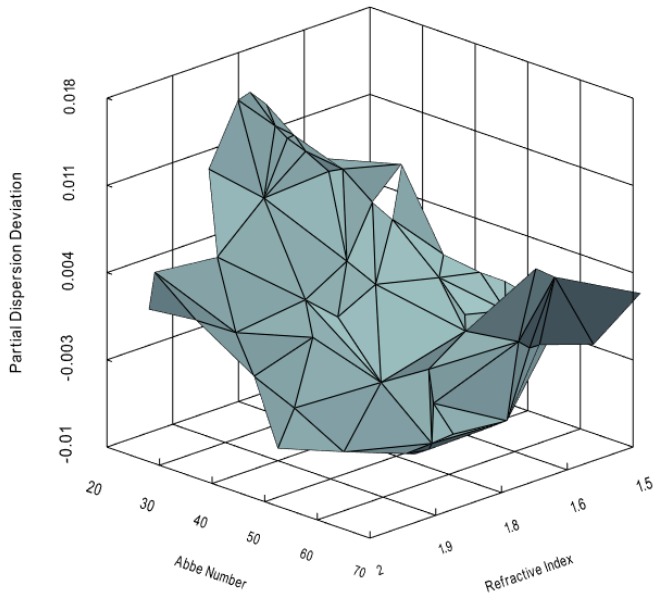


Figure S6: Glass catalog mesh for the 3-dimensional case. We show the mesh for the MOL study of Section S5, where 104 recommended glasses from the Ohara catalog are modeled. Vertices correspond to the catalog entries, and edges represent the Delaunay tessellation. The mesh representation closely matches the outside shape of the glass catalog, therefore discouraging the optimization process from exploring empty space.

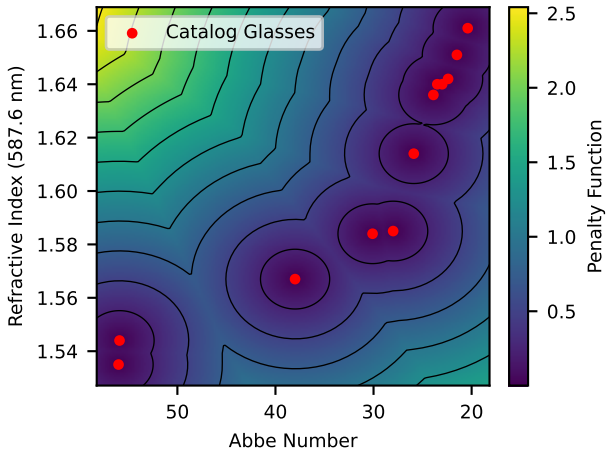


Figure S7: Calculation of the glass catalog distance residuals. The color indicates the distance to the nearest entry in the glass catalog.

For each set of glass variables $\mathbf{g}$, the distance to the closest catalog entry is

$$L_{\text{GCD}}(\mathbf{g}) = \min_{\mathbf{g}' \in \mathcal{G}} \left\| \mathbf{g} - \mathbf{g}' \right\| , \tag{S56}$$

where $\mathcal{G}$ is the set of all glass variables in the catalog. The residual vector for n_g glass variables is then:

$$\boldsymbol{\ell}_{\text{GCD}} = \left[L_{\text{GCD}}(\mathbf{g}_1), L_{\text{GCD}}(\mathbf{g}_2), \ldots, L_{\text{GCD}}(\mathbf{g}_{n_g}) \right] . \tag{S57}$$

In practice, the weights of the glass catalog distance residuals are typically set lower than those of the glass mesh distance residuals, allowing flexibility to explore the design space. As optimization progresses, the weights of the catalog residuals are gradually increased following a piecewise exponential schedule. This approach ensures that the final design conforms to real catalog materials while preserving the freedom to explore unconventional solutions during the early stages.

S2.3 Optimization Algorithms

In this section, we describe two families of optimization algorithms for lens design. Gradient descent-based methods, detailed in Section S2.3.1, rely on the gradient $\nabla_{\boldsymbol{\theta}} L$ of a scalar objective function. In contrast, Gauss-Newton-based methods, including the LM algorithm (Section S2.3.2), use the Jacobian $\mathbf{J} \in \mathbb{R}^{M \times N}$, where $\mathbf{J}_{mn} = \partial \ell_m / \partial \theta_n$.

As we expect the number of residuals M to exceed the number of variables N, we compute $\mathbf{J}$ using forward-mode automatic differentiation, which scales with N rather than M.

S2.3.1 Gradient Descent Methods

The basic step in gradient descent algorithms is

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k - \alpha_k \nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}_k) , \tag{S58}$$

where α_k is the learning rate at iteration k. Variants like momentum [20] and adaptive learning rates [25] improve upon plain gradient descent, giving rise to popular algorithms like Adam [13]. Machine learning often uses the stochastic gradient descent (SGD) variant [2], which computes gradients on random data subsets.

For a sum-of-squares objective $L(\boldsymbol{\theta})$ from Equation (S27), the gradient is

$$\nabla_{\boldsymbol{\theta}} L = \mathbf{J}^T \boldsymbol{\ell} . \tag{S59}$$

Unlike methods leveraging the Jacobian $\mathbf{J}$, gradient descent uses only the gradient vector of size N, losing individual residual ℓ_m information.

S2.3.2 Levenberg-Marquardt Algorithm

The LM algorithm refines the Gauss-Newton method to improve stability in non-linear least-squares problems. Assuming the vector objective $\boldsymbol{\ell}$ is locally linear, $\boldsymbol{\ell} \approx \boldsymbol{\ell}_0 + \mathbf{J}\boldsymbol{\theta}$, the sum-of-squares objective from Equation (S27) simplifies to

$$L(\boldsymbol{\theta}) \approx \frac{1}{2} \|\boldsymbol{\ell}_0 + \mathbf{J}\boldsymbol{\theta}\|^2 = \frac{1}{2} \|\mathbf{J}\boldsymbol{\theta}\|^2 + \boldsymbol{\theta}^T \mathbf{J}^T \boldsymbol{\ell}_0 + \frac{1}{2} \|\ell_0\|^2 \,. \tag{S60}$$

Minimizing this quadratic approximation yields the Gauss-Newton solution:

$$\boldsymbol{\theta}^* = -\left(\mathbf{J}^T \mathbf{J}\right)^{-1} \mathbf{J}^T \boldsymbol{\ell}_0 \,. \tag{S61}$$

However, this solution is unreliable when the Jacobian matrix is ill-conditioned [21]. In practice, the Gauss-Newton method is a poor fit for optical design because of the instability caused when the Jacobian is rank-deficient or nearly so. To address this, the LM algorithm [18, 19] introduces a damping term, solving instead:

$$\boldsymbol{\theta}^* = -\left(\mathbf{J}^T \mathbf{J} + \lambda \mathbf{D}^2\right)^{-1} \mathbf{J}^T \boldsymbol{\ell}_0 \,, \tag{S62}$$

where λ is the damping factor, and $\mathbf{D}^2$ is the damping matrix, commonly set to the diagonal of $\mathbf{J}^T\mathbf{J}$ to preserve scale invariance [21]. Equation (S62) is the solution to the modified linear least-squares problem

$$L(\boldsymbol{\theta}) = \frac{1}{2} \|\boldsymbol{\ell}_0 + \mathbf{J}\boldsymbol{\theta}\|^2 + \frac{1}{2}\lambda \|\mathbf{D}\boldsymbol{\theta}\|^2 \,. \tag{S63}$$

As $\lambda \to 0$, the solution approaches Gauss-Newton; when $\lambda \gg \left\|\mathbf{J}^T\mathbf{J}\right\|$, it resembles gradient descent, effectively interpolating between the two methods.

Several strategies are used in practice to improve on the basic LM algorithm. Below, we list some strategies employed in our implementation. We summarize the user-defined parameters for the optimizer in Table S7.

Damping matrix The damping matrix $\mathbf{D}$, often initialized to the diagonal of $\mathbf{J}^T\mathbf{J}$, preserves scale invariance but may suffer from parameter evaporation [33], where the loss is nearly insensitive to some parameters, as reflected by small diagonal entries in $\mathbf{D}$ In this instance, the LM algorithm may take exceedingly large steps for these parameters, driving them to unreasonably high values and causing instability. To mitigate this, a running maximum strategy is applied to stabilize the damping matrix over iterations:

$$\mathbf{D}_{k+1} = \beta \max\left(\mathbf{D}_k, \operatorname{diag}(\mathbf{J}_k^T \mathbf{J}_k)^{1/2}\right) + (1-\beta)\operatorname{diag}(\mathbf{J}_k^T \mathbf{J}_k)^{1/2} \,, \tag{S64}$$

Table S7: Summary of user-defined parameters for the LM optimizer, along with sensible default values.

Parameter	Description
$\lambda_0 = 1$	Initial damping factor
$\epsilon = 1 \times 10^{-6}$	Minimum value for the entries of the damping matrix at the beginning of the optimization process
$\beta = 0.99$	Running maximum parameter for damping matrix
IF $= 2$, DF $= 3$	Increase and decrease factors for damping factor λ; if the loss increases or decreases, the damping factor is scaled up or down by these parameters, respectively
TF $= 1$	Tolerance factor for rejecting updates; when the loss increases beyond this threshold, the step is rejected

where $\beta \in (0, 1)$ is a user-defined parameter controlling the weight of past values. Additionally, we initialize the damping matrix to

$$\mathbf{D}_0 = \max\left(\text{diag}(\mathbf{J}_0^T \mathbf{J}_0)^{1/2}, \epsilon\right) , \tag{S65}$$

with ϵ as the user-defined initial minimum value.

Damping factor The damping factor λ is updated adaptively during optimization. When the loss decreases, the linear approximation to the objective function is reasonable, so λ is decreased to favor Gauss-Newton behavior. Conversely, when the loss increases, λ is raised to prioritize the stability of gradient descent. A common strategy, as proposed by Transtrum and Sethna [33], is to increase λ by a factor of 2 (increase factor, IF) when the loss increases and to decrease it by a factor of 3 (decrease factor, DF) when the loss decreases.

Rejecting updates To further ensure stability, the algorithm employs a mechanism to reject updates that increase the loss excessively. Steps are rejected if the loss grows by more than a user-defined tolerance factor TF. A typical default value is TF $= 1$, which only accepts steps that reduce the loss.

S3 End-to-End Optimization

In this section, we provide additional information on our end-to-end (E2E) optimization strategy, applied to E2E image restoration for optimizing a lens design, an image restoration model (IRM), or both components jointly. We build on the optical modeling operations presented in Section S1 and the lens optimization strategy presented in Section S2.

In Section S3.1, we provide details on the formulation of the generalized transverse ray aberrations (GTRA) as an approach to task-driven optimization of optical systems. In Section S3.2, we discuss models for image restoration, and describe our single-stage conditioned IRM that combines Wiener deconvolution with data-driven approaches. In Section S3.3, we present the E2E optimization strategy that jointly optimizes the lens and IRM parameters.

S3.1 Generalized Transverse Ray Aberrations

We provide further technical details on our derivation of the *generalized transverse ray aberrations* (GTRA) objective as an approach to task-driven optimization of optical systems, complementing the formulation discussed in the main paper.

We start from the key assumption that the scalar loss function $L_{\text{TD}}(\boldsymbol{\theta})$ is fully determined by the spot diagrams $\boldsymbol{\epsilon}(\boldsymbol{\theta})$ as a representation of the lens. We note that the objective can also depend on other variables quantities, such as the input image $\mathbf{I}$ or the IRM parameters $\boldsymbol{\phi}$ in the case of E2E image restoration, which we treat as constants for a given step.

Let $\boldsymbol{\epsilon}_0$ denote the spot diagrams at the current step, $L_{\boldsymbol{\epsilon}_0}$ the scalar loss, and $\nabla L_{\boldsymbol{\epsilon}_0}$ its gradient, computed using backward-mode automatic differentiation. From these quantities, we convert the scalar loss into a sum-of-squares objective after two successive approximations,

$$
\begin{aligned}
L_{\text{TD}}(\boldsymbol{\theta}) &\approx L_{\boldsymbol{\epsilon}_0} + \nabla L_{\boldsymbol{\epsilon}_0}^T \left(\boldsymbol{\epsilon}(\boldsymbol{\theta}) - \boldsymbol{\epsilon}_0\right) \\
&= L_{\boldsymbol{\epsilon}_0} \left(1 + 2\frac{\nabla L_{\boldsymbol{\epsilon}_0}^T \left(\boldsymbol{\epsilon}(\boldsymbol{\theta}) - \boldsymbol{\epsilon}_0\right)}{2L_{\boldsymbol{\epsilon}_0}}\right) \\
&= L_{\boldsymbol{\epsilon}_0} \frac{\nabla L_{\boldsymbol{\epsilon}_0}^T}{\|\nabla L_{\boldsymbol{\epsilon}_0}\|} \left(\frac{\nabla L_{\boldsymbol{\epsilon}_0}}{\|\nabla L_{\boldsymbol{\epsilon}_0}\|} + 2\frac{\|\nabla L_{\boldsymbol{\epsilon}_0}\| \left(\boldsymbol{\epsilon}(\boldsymbol{\theta}) - \boldsymbol{\epsilon}_0\right)}{2L_{\boldsymbol{\epsilon}_0}}\right) \\
&\approx L_{\boldsymbol{\epsilon}_0} \left\| \frac{\nabla L_{\boldsymbol{\epsilon}_0}}{\|\nabla L_{\boldsymbol{\epsilon}_0}\|} + \frac{\|\nabla L_{\boldsymbol{\epsilon}_0}\| \left(\boldsymbol{\epsilon}(\boldsymbol{\theta}) - \boldsymbol{\epsilon}_0\right)}{2L_{\boldsymbol{\epsilon}_0}} \right\|^2 \\
&= \underbrace{\frac{1}{2} \frac{\|\nabla L_{\boldsymbol{\epsilon}_0}\|^2}{2L_{\boldsymbol{\epsilon}_0}}}_{w} \left\| \boldsymbol{\epsilon}(\boldsymbol{\theta}) - \underbrace{\left(\boldsymbol{\epsilon}_0 - 2\frac{L_{\boldsymbol{\epsilon}_0} \nabla L_{\boldsymbol{\epsilon}_0}}{\|\nabla L_{\boldsymbol{\epsilon}_0}\|^2}\right)}_{\boldsymbol{\epsilon}'} \right\|^2 .
\end{aligned}
\tag{S66}
$$

The first approximation is a linearization of the scalar loss function relative to the spot diagrams $\boldsymbol{\epsilon}$, similar to the derivation of the Levenberg-Marquardt (LM) update equation. The second approximation is justified by the term in $\boldsymbol{\epsilon}(\boldsymbol{\theta}) - \boldsymbol{\epsilon}_0$ becoming increasingly smaller when close to

convergence[1]. Equation (S66) has the form of a least-squares optimization objective

$$L_{\mathrm{TD}}(\boldsymbol{\theta}) \approx \frac{1}{2} \left\| \boldsymbol{\ell}_{\mathrm{GTRA}}(\boldsymbol{\theta}) \right\|^2 , \tag{S67}$$

where

$$\boldsymbol{\ell}_{\mathrm{GTRA}}(\boldsymbol{\theta}) = \sqrt{w} \left(\boldsymbol{\epsilon}(\boldsymbol{\theta}) - \boldsymbol{\epsilon}' \right) . \tag{S68}$$

Unlike the spot-optimized scenario, the control values $\boldsymbol{\epsilon}'$ in GTRA do not correspond to the spot centroid, but vary individually for each ray according to the current value of the scalar loss function $L_{\boldsymbol{\epsilon}_0}$ and its gradient $\nabla L_{\boldsymbol{\epsilon}_0}$.

In practice, the optimization objective includes additional terms that enforce various constraints on the lens in addition to the task-driven loss function. We construct the complete sum-of-squares objective by concatenating the residuals relative to these constraints with $\boldsymbol{\ell}_{\mathrm{GTRA}}(\boldsymbol{\theta})$.

Bounding control values The finite size of the virtual sensor grid used for point spread function (PSF) estimation can result in control values $\boldsymbol{\epsilon}'$ exceeding the grid boundaries, leading to numerical instability as rays outside these limits are excluded from the optimization process. To address this, we clip the control values $\boldsymbol{\epsilon}'$ to the grid boundaries and rescale them to preserve the scalar value of the sum-of-squares objective. Where computationally feasible, a sufficiently large virtual sensor grid should be selected to minimize the occurrence of this issue.

Jacobian matrix Computing LM updates, as shown in Equation (S62), requires the loss vector $\boldsymbol{\ell}(\boldsymbol{\theta})$ and its Jacobian matrix $\mathbf{J}$ with respect to the lens parameters $\boldsymbol{\theta}$. The Jacobian is computed using forward-mode automatic differentiation, consistent with the lens optimization strategy in Section S2. Since the downstream loss vector $\boldsymbol{\ell}_{\mathrm{GTRA}}(\boldsymbol{\theta})$ from Equation (S68) depends only on the spot diagrams, forward-mode differentiation is applied exclusively to the ray-tracing operations that generate spot diagrams. For the remainder of the pipeline, including image simulation and restoration, forward-mode differentiation is unnecessary, and only backward-mode differentiation is used to compute $\nabla L_{\boldsymbol{\epsilon}_0}$, which determines the weight w and control values $\boldsymbol{\epsilon}'$. This approach ensures the computation of the Jacobian matrix remains as efficient as in standard lens design optimization.

S3.2 Image Restoration Models

Image restoration enhances image quality by removing degradations like noise, blur, and artifacts. In E2E design for computer vision systems, it serves two key functions. First, it compensates for degradations from the optical system, improving visual quality during deployment. Second, it

[1]We use the following approximation, which holds when $\|\boldsymbol{v}\| \ll \|\boldsymbol{u}\|$:

$$\begin{aligned} \|\boldsymbol{u}+\boldsymbol{v}\|^2 &= \boldsymbol{u}^T\boldsymbol{u} + 2\boldsymbol{u}^T\boldsymbol{v} + \boldsymbol{v}^T\boldsymbol{v} \\ &\approx \boldsymbol{u}^T\boldsymbol{u} + 2\boldsymbol{u}^T\boldsymbol{v} \\ &= \boldsymbol{u}^T(\boldsymbol{u}+2\boldsymbol{v}) \end{aligned}$$

provides a more effective loss function for optical system design, guiding lens optimization to maximize actual image quality rather than relying on intermediate metrics like modulation transfer function (MTF) or spot size. While the first function applies in post-production, the second is crucial during the design and optimization phase.

Combining classical and data-driven methods Research indicates that combining classical methods, such as Wiener deconvolution, with deep learning can improve restoration quality [8, 9, 12, 34]. Conditioned IRMs offer an effective approach by integrating physical degradation models with data-driven learning. These models are less prone to overfitting compared to purely data-driven approaches. Additionally, this hybrid approach results in more lightweight models, as the classical method (e.g., Wiener deconvolution) handles the inversion of optical degradations, leaving the neural network to focus on learning only the residual errors.

Conditioned IRMs in E2E optical design Conditioned IRMs are especially valuable in E2E optical design, where the optical system's response is well-understood and can be accurately modeled. A natural choice for conditioning the neural network is to provide the system's PSFs as input. Although PSF vary spatially, they can be approximated as constant over small image patches, enabling a patch-based approach for effective image restoration.

Existing methods for conditioning networks Several methods for conditioning neural networks with PSFs have been proposed. In [8, 34], a multi-scale architecture is proposed, where scaled PSFs are input at each scale, performing Wiener deconvolution and fusing the results for the final image. In contrast, [9, 15] frame the problem as an iterative optimization process using half-quadratic splitting.

Proposed conditioned IRM We propose a simpler, single-stage conditioned IRM. In this model, Wiener deconvolution is performed once and is both preceded and followed by separate UNet-like [23] models, with NAFNet [3] for the UNet architecture. This approach streamlines the process, making for an effective and lightweight model for benchmarking our E2E optimization strategy. Notably, the numerous skip connections of NAFNet allow our conditioned model to behave close to an identity mapping in the beginning of the E2E process, which is key in stabilizing the E2E lens optimization. In addition, the use of Wiener deconvolution facilitates quick adaptation to the optical system's response. In practice, Wiener deconvolution requires an estimate of the image's signal-to-noise ratio (SNR). In our experiments, we treat the SNR as an optimizable parameter, allowing it to be learned and fine-tuned during the image restoration process.

S3.3 End-to-End Optimization Procedure

During joint training, the IRM adapts to the lens, while the lens is optimized to capture the most important features of the virtual scene for the IRM. This iterative process enables the lens and IRM to co-adapt, resulting in an imaging system that can maximize perceptual image quality.

IRM and lens initialization In joint training, the IRM can be trained from scratch, particularly when its architecture allows for rapidly matching and enhancing the raw image quality, as in our proposed conditioned IRM model. However, the lens must already be in a viable state before joint training begins to minimize ray failures and ensure that the PSF arrays collect most of the incoming energy.

Adapting the LM algorithm for E2E optimization As discussed in Section S2, lens optimization involves three critical and largely independent aspects: lens parameterization, optimization objective, and optimization algorithm. The parameterization and optimization strategy remain consistent with the standard lens optimization approach outlined in Section S2. However, the optimization objective is adapted to minimize the downstream loss function, which is then converted into the GTRA objective as detailed in Section S3.1. The Levenberg-Marquardt (LM) algorithm, typically used for static optimization objectives, is well-suited for the lens optimization step. However, in the context of our E2E optimization strategy, the objective function changes at each step due to the random selection of image patches and ongoing updates to the IRM parameters. Despite these changes, we find that the LM algorithm remains robust, provided the loss function does not fluctuate drastically between optimization steps. To stabilize the optimization, we employ a pseudo-random sampling strategy, as described in Section S3.3.1, which ensures balanced representation of all sensor regions in each batch.

S3.3.1 Joint Optimization Step

Our proposed joint optimization process consists of a two-step optimization procedure for each training iteration. First, a minibatch of image patches and sensor regions is sampled. In the first step, we perform an E2E optimization step on the lens parameters using the LM algorithm, while freezing the IRM parameters. In the second step, with the same image patches and sensor regions, we perform an optimization step on the IRM parameters using stochastic gradient descent (SGD) optimizers such as Adam [13], while freezing the lens parameters. These steps are iterated for a fixed number of training steps, with learning rate scheduling applied to the IRM optimizer.

Here, we detail all operations required in an optimization step for either the lens or the IRM, and highlight the key differences between the two steps where appropriate. Our methodology borrows from standard practice for training IRMs.

Sampling of image patches To compute the loss function, we first form a batch of image patches from the training dataset. While IRMs can support full-resolution images, training on large images can be computationally expensive and memory intensive. Furthermore, with conditioned IRMs, the PSFs are typically constant over small patches of the image, making patch-based training a natural choice. Therefore, we start the training step by sampling a batch of image patches from the training dataset with, e.g., a batch size of 16 and a standard patch size of 256×256 pixels. Data augmentation techniques can be applied at this step to increase the diversity of the training data, like random cropping and flipping.

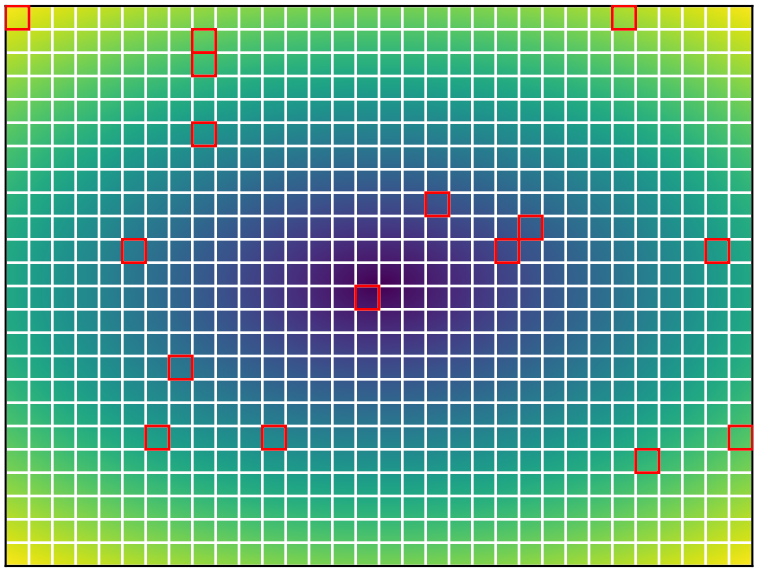


Figure S8: Strategy for sampling sensor regions during training. The sensor resolution is divided into a grid of non-overlapping tiles, each representing a distinct sensor region. A color map visualizes the normalized coordinates, with blue indicating the center of the field of view (FOV) and yellow marking the corners. For each image patch in the minibatch, a sensor region is randomly selected, as illustrated by the red squares for a minibatch of 16 patches. The selected region represents where the image patch is captured, guiding how imaging simulation is performed. This ensures diverse optical field representation within each minibatch. In this example, a sensor with a resolution of 8192×6144 pixels is divided into a grid of 32×24 tiles, each of size 256×256 pixels.

Sensor region sampling strategy In typical optical systems, the PSFs vary significantly across the sensor, which must be accounted for in the simulation. To ensure all sensor regions are equally represented in the training data, we divide the sensor into a grid of non-overlapping tiles, as shown in Figure S8. For each image patch in the minibatch, a sensor region corresponding to the where the image patch is captured on the sensor is randomly selected. To prevent instability in lens optimization, we use a pseudo-random sampling strategy. All tiles are sorted by their distance from the sensor's center, and then pseudo-randomly sampled to ensure a balanced distribution across the sensor for each minibatch. This strategy helps maintain consistency between iterations, improving the robustness of lens optimization.

Computation of spot diagrams Before the imaging simulation, spot diagrams are computed for various field values and wavelengths (see Section S1.3.3). In lens optimization only, forward-mode automatic differentiation is used alongside the computation of spot diagrams and other soft constraints, enabling the calculation of the Jacobian matrix.

Image formation simulation Each selected sensor region is further divided into a grid of sub-tiles, e.g., 3×3 sub-tiles, and we evaluate the PSFs at the center of each sub-tile. We apply a spatially varying convolution between the image patch $\mathbf{I}$ and corresponding PSFs to simulate the optical system's response to the image patch (see Section S1.5).

Noise simulation Additive white Gaussian noise is added to the image during training to approximate noise in real-world scenarios. We model the noise with a standard deviation of, e.g., 0.5 % of the maximum pixel value. Adding noise is crucial for regularizing the IRM training, preventing the model from overfitting to the ideal optical response and ensuring that it generalizes well to real-world data. The resulting aberrated image patch is labeled as $\mathbf{I}'$.

Image restoration The degraded image patch $\mathbf{I}'$ is restored using the conditioned IRM. The IRM takes the degraded patch and, optionally, the PSFs, as input and outputs the restored patch $\mathbf{I}''$. In the case of a conditioned IRM, the PSFs are provided as input to the model to guide the restoration process. For conditioned IRMs, the PSFs for all sub-tiles are averaged together to form a single input PSF. This prevents overfitting to specific PSFs used in the simulation.

Downstream loss computation The downstream loss L_{E2E} is computed as the average discrepancy between batches of images $\mathbf{I}$ and $\mathbf{I}''$. In our experiments, we use the mean absolute error (MAE) loss for the IRM training, following standard practice in image restoration tasks. For lens optimization, we use the mean squared error (MSE) loss, which empirically provides better stability and robustness.

Automatic differentiation For IRM optimization, only backward-mode automatic differentiation is used to compute gradients of the downstream loss with respect to the IRM parameters, as the lens parameters are frozen. In lens optimization, backward-mode automatic differentiation is applied through the imaging simulation and image restoration to compute gradients with respect to the spot diagrams, which are required to compute the GTRA objective.

[Lens optimization only] Sum-of-squares objective During lens optimization, the downstream loss is converted into GTRA (see Section S3.1). The least-squares objective includes the GTRA term, along with additional terms that enforce soft constraints on the lens (see Section S2.2).

Optimization In lens optimization, the LM algorithm is employed to optimize the lens parameters (see Section S2.3.2). After each optimization step, the sum-of-squares objective is reevaluated. Based on the change in the objective function, the damping factor is adjusted to favor convergence, and the rejection mechanism is applied to reject the step if necessary, which is crucial in preventing divergence. In IRM optimization, the parameters are updated using SGD optimizers such as Adam [13]. Optimization terminates after a predefined number of steps.

S4 Additional Results for Smartphone Telephoto Lenses

In this section, we provide additional information on the study on the end-to-end (E2E) design of smartphone telephoto lenses.

In Section S4.1, we detail the problem formulation and baseline design for smartphone telephoto lenses. In Section S4.2, we select multiple configurations of interest and compare their optical performance in terms of effective spot radius (ESR). In Section S4.3, we investigate the image quality of baseline lenses and compare them with and without image restoration. In Section S4.4, we provide additional analysis and validation on E2E-optimized imaging systems.

S4.1 Problem Formulation and Baseline Design

In this section, we formulate the problem of designing smartphone telephoto lenses and introduce a baseline design. Telephoto lenses are characterized by a total track length (TTL) shorter than their effective focal length (EFL), quantified by the telephoto ratio (TR), defined as $\mathrm{TR} = \mathrm{TTL}/\mathrm{EFL}$. Typical TRs range from 0.6 to 0.85 [30], with lower ratios posing greater challenges for aberration correction.

S4.1.1 Specifications

For the purpose of this study, we consider the specifications listed in Table S8.

The target sensor (Samsung ISOCELL GN5) features a resolution of 48 MP with a pixel pitch of 1 µm. For image simulation, patches of 256×256 px are used, yielding 768 (32×24) discrete sensor regions without overlap, as outlined in Section S3.3.1.

The optical design is guided by several nominal parameters: an image height of 5.1 mm (half the sensor diagonal), a target EFL of 25.9 mm, and a target f-number of 2.8, which result in a field of view (FOV) of 22.3°. However, the actual EFL and f-number may vary slightly during optimization. To maintain the paraxial image height rather than strictly fixing the EFL, the last curvature of the system is solved according to the image-height-consistency focal length, as detailed in Section S2.1.3. Similarly, the entrance pupil diameter (EPD), rather than the f-number, is fixed at 9.25 mm to prevent the optimization process from reducing the EFL to shrink the aperture size, which could simplify aberration correction undesirably.

Performance and manufacturability are addressed through soft constraints in the least-squares optimization objective, as further detailed in Section S4.1.4. As our imaging simulation does not model distortion and relative illumination, we include constraints to address these imaging performance criteria.

S4.1.2 Visible Spectrum

In Figure S9, we illustrate the spectrum used for designing and evaluating smartphone telephoto lenses. We start with a representative spectrum for smartphone imaging, consisting of a regular

Table S8: Summary of specifications for the telephoto lens study.

Parameter	Value
Sensor specifications	
Sensor	1/1.5″ Samsung ISOCELL GN5
Sensor resolution	8160 × 6144 px (we use 8192 × 6144 px to simplify)
Pixel pitch	1 µm
Sensor diagonal	10.24 mm
First-order specifications	
Field of view	22.3°
Nominal image height	5.1 mm
Nominal EFL	25.9 mm
Nominal f-number	2.8
Entrance pupil diameter	9.25 mm
Nominal spectrum	Detailed in Section S4.1.2
Image quality specifications	
Distortion	$\leq$ 2 % across the field
Relative illumination	$\geq$ 70 % across the field
Manufacturability specifications	
Element thickness	$\geq$ 0.25 mm
Thickness to center thickness ratio	1/3 $\leq$ ratio $\leq$ 3 across element diameter
Air gap spacing	$\geq$ 0.1 mm
Mechanical flange focal length	$\geq$ 5.1 mm
Surface normals of refractive elements	$\leq$ 30°
Angles of incidence/refraction	$\leq$ 60°
Miscellaneous specifications	
Infrared (IR) filter	BK7 with thickness 0.11 mm
Refractive materials	Detailed in Section S4.1.3

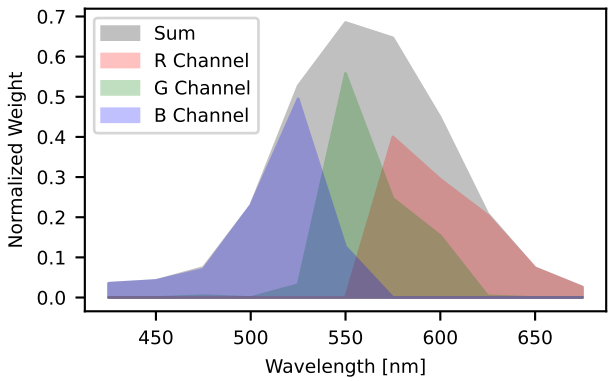


Figure S9: Spectrum used for designing and evaluating smartphone telephoto lenses, with 11 wavelengths from 425–675 nm. The spectrum is split into RGB channels for imaging simulation and E2E design.

grid of 11 wavelengths ranging from 425–675 nm, with an average wavelength of 558.7 nm and a standard deviation of 44.5 nm. This monochromatic spectrum is used for spot size optimization.

For imaging simulation and task-driven design, we split the spectrum evenly into RGB channels, ensuring that their sum matches the original spectrum. First, all weight for 425 nm is assigned to the B channel, and all weight for 675 nm is assigned to the R channel. An optimization strategy is then applied to minimize the average extent of the channels. The weights for the RGB channels are listed in Table S9.

Table S9: Normalized weights for the RGB channels of the telephoto lens, corresponding to the weight matrix $\mathbf{W}_w$ in Section S1.5.1, used to combine individual wavelength PSFs into RGB PSFs. The wavelength range from 500–625 nm accounts for 91.6 % of the spectrum.

Wavelength (nm)	R channel	G channel	B channel	Average
425	0.0000	0.0000	0.0355	0.0119
450	0.0000	0.0000	0.0425	0.0142
475	0.0000	0.0046	0.0703	0.0250
500	0.0000	0.0000	0.2299	0.0766
525	0.0000	0.0321	0.4951	0.1757
550	0.0000	0.5584	0.1267	0.2284
575	0.4002	0.2467	0.0000	0.2156
600	0.2946	0.1541	0.0000	0.1495
625	0.2052	0.0041	0.0000	0.0698
650	0.0739	0.0000	0.0000	0.0246
675	0.0261	0.0000	0.0000	0.0087

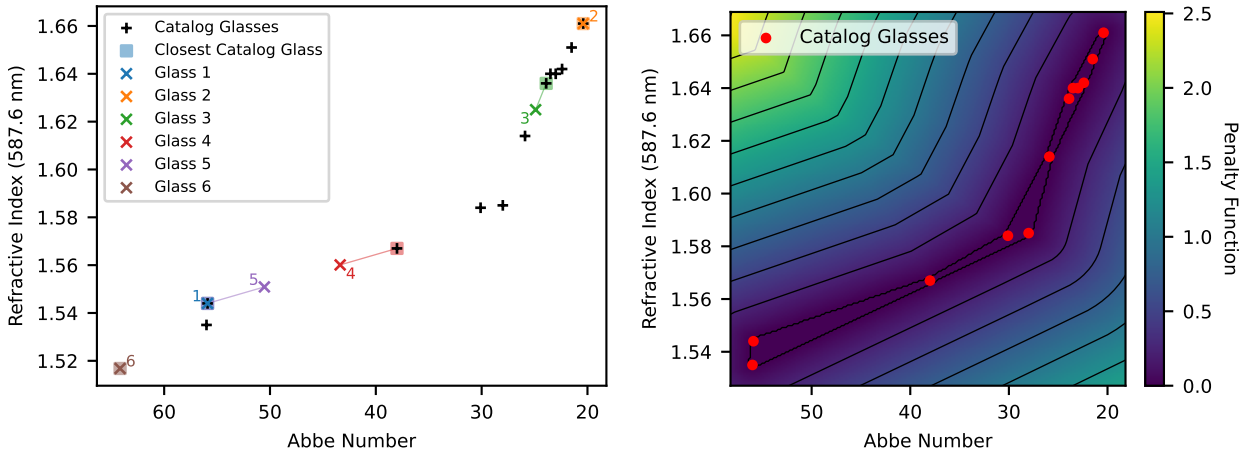
S4.1.3 Refractive Materials

For smartphone telephoto lenses, moldable optical plastics are the primary material choice. Table S10 lists the refractive index and Abbe number of materials from a curated catalog. As shown in Figure S10b, these materials have limited dispersion properties, mainly concentrated along a path where the Abbe number decreases as the refractive index increases.

Our optimization strategy allows the refractive index and Abbe number to vary continuously, as illustrated in Figure S10a, but each material is bound to its closest catalog counterpart at the midpoint of the optimization. Once bound, the glass variables are removed from the optimization process, the algorithm is reset to its original parameters, and optimization continues. Only in this study, the materials are optimized when the lens is optimized for spot size, then frozen for the E2E optimization process. Figure S10b illustrates the penalty function used for the catalog of moldable materials, based on the distance between the optimized materials and the 2D catalog mesh, as detailed in Section S2.2.4.

S4.1.4 Constraints and Optimization Objective

Based on the specifications in Table S8, we define the constraints and optimization objective for the telephoto lens design. A summary is provided in Table S11.



(a) Glass Optimization

(b) Glass Penalty

Figure S10: Optimization of refractive materials for the telephoto lens. (a) Refractive material variables are free to vary during the first half of optimization, after which they are bound to the closest catalog counterpart. (b) The penalty function ensures the optimized materials remain close to the catalog materials.

Table S10: Moldable refractive materials for the telephoto lens, sorted by increasing refractive index. The Abbe number decreases monotonically as the refractive index increases.

Material	Refractive index (587.6 nm)	Abbe number
K26R	1.535	56.0
5514ML	1.544	55.9
EP3500	1.567	38.0
PC(BLK)	1.584	30.1
SP5580	1.585	28.0
EP4500	1.614	25.9
EP5000	1.636	23.9
EP6000	1.640	23.5
SP3810	1.640	23.0
OKP1	1.642	22.4
EP7000	1.651	21.5
EP8000	1.661	20.4

Table S11: Summary of constraints and optimization objective for the telephoto lens design problem. In our experiments, we note that either ℓ_{TRA} or ℓ_{GTRA} is used at a given time, whether the lens is spot-optimized or E2E-optimized. All other residuals are used in all experiments with the same weights except where mentioned.

Residuals	Description	Notes
ℓ_{TRA}	Transverse ray aberrations residuals. Similar to optimizing the spot size. See Section S2.2.1.	Only used when optimizing lenses for spot radius—labeled in this chapter as *spot-optimized* lenses.
ℓ_{GTRA}	Generalized transverse ray aberrations residuals. Used to convert the scalar image quality loss into a least-squares objective. See Section S3.1.	Only for image-driven optimization—labeled as *E2E-optimized* lenses.
ℓ_{RP}	Ray path residuals. Uses traced rays as probes to ensure that a minimum and maximum horizontal distance are met between pairs of adjacent optical surfaces. See Section S2.2.2.	In refractive elements (minus the infrared (IR) filter), the lower threshold is set to the highest value between 0.25 mm or CT/3, where CT is the center thickness. The upper threshold is set to the lowest value between 4 mm or 3CT. In air gaps, a lower threshold is set to 0.1 mm in general and 5.1 mm for the gap between the last optical surface and sensor.
ℓ_{RA}	Ray angle residuals. Limits angles of refraction and incidence within a given range. See Section S2.2.2.	The threshold is set to $\pm 60^\circ$.
ℓ_{SN}	Surface normal residuals. Limits the angle between the normals of the refractive surfaces and the optical axis. See Section S2.2.2.	The threshold is set to 30°.
ℓ_{D}	Distortion residuals. Limits distortion to a given range over the optical field. See Section S2.2.3.	The threshold is set to $\pm 2\,\%$.
ℓ_{RI}	Relative illumination residuals. Enforces relative illumination above a given threshold over the optical field. See Section S2.2.3.	The minimum threshold is set to 70 %.
ℓ_{GMD}	Glass mesh distance residuals. Ensures that the optimized refractive index and Abbe number are realistic relative to the glass catalog. Only used in spot size optimization. See Section S2.2.4.	The maximum edge threshold is set to 1.25.
ℓ_{GCD}	Glass catalog distance residuals. Forces the glass variables to commit to a catalog entry. Only used in spot size optimization. See Section S2.2.4.	

In the following experiments, we investigate two settings: lenses optimized for spot size (*spot-optimized lenses*) and lenses optimized in an E2E fashion (*E2E-optimized lenses*). For spot-optimized lenses, the optimization minimizes the transverse ray aberrations (TRA), equivalent to optimizing spot size. For E2E-optimized lenses, the goal is to minimize the image disparity loss between virtual scenes and restored images, with the loss converted into generalized transverse ray aberrations (GTRA), as detailed in Section S3.1. Both settings use the same optimization objective, differing only in the residuals ℓ_{TRA} and ℓ_{GTRA}.

The primary residuals driving the optimization are ℓ_{TRA} and ℓ_{GTRA}, which serve as the main optimization targets. All other residuals are soft constraints, with high weights to ensure enforcement. Most constraints ($\ell_{\text{RP}}, \ell_{\text{RA}}, \ell_{\text{SN}}, \ell_{\text{GP}}, \ell_{\text{GD}}$) are manufacturability constraints, while others ($\ell_{\text{D}}, \ell_{\text{RI}}$) focus on image quality, addressing aspects not already covered by ℓ_{TRA} or ℓ_{GTRA}.

S4.1.5 Baseline Design

In this section, we present an example design for smartphone telephoto lenses, shown in Figure S11, and detail our design choices. The design was optimized for spot size with the objective outlined in Section S4.1.4, resulting in a TR of 0.8. It consists of five aspheric refractive elements and an infrared (IR) filter, and serves as a representative model for all lenses in this study.

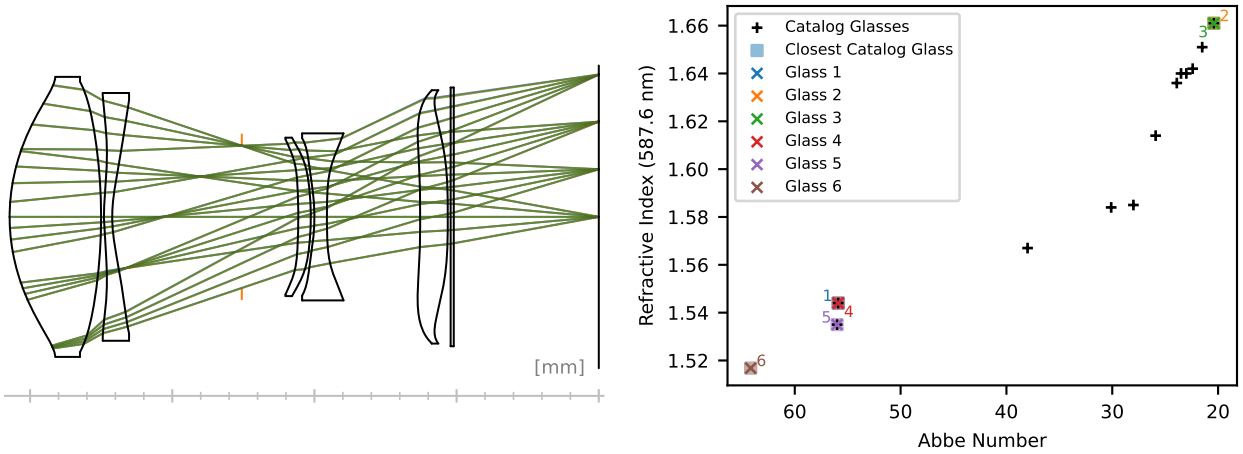
Front aperture To minimize the TR, some vignetting is necessary. We place a front aperture in object space, in front of the first refractive element, with a diameter of 9.25 mm, equal to the EPD. Its location is fixed at 1 mm to the right of the vertex of the first optical surface. This configuration balances relative illumination with optical performance, as moving the stop deeper would improve spot size at the cost of relative illumination. To maintain illumination above the desired threshold across the field, we include a soft constraint for relative illumination.

Aperture stop The aperture stop is located between the second and third elements, delineating the front and rear groups of the design. While placing the stop between the third and fourth elements may slightly improve performance for small TR values, this effect is minimal and does not justify additional lens configurations in this study.

IR filter An IR filter with a thickness of 0.11 mm and N-BK7 material is placed at the rear of the lens. The filter's spacing, material, and surface curvatures are not optimized.

Solves Two solves are used in all experiments: one for the last optimizable surface curvature to enforce the desired paraxial image height at the sensor, and another for the last airspace to enforce the target TTL. These two variables are removed from the optimization process.

Optimizable variables The lens has 90 optimizable variables when materials are optimized and 80 when materials are fixed. This includes 11 spacings (out of 13), 9 curvatures (out of 12), and 10×6 aspheric coefficients.



(a) 2D Layout

(b) Refractive Materials

Figure S11: Baseline design for the telephoto lens with a TR of 0.8. (a) 2D layout of the lens, composed of five aspheric refractive elements and an infrared (IR) filter. The front aperture is placed in object space, in front of the first element, and the aperture stop is positioned between the first two elements and the last three. (b) Refractive materials for the lens. At this TR, the first, fourth, and fifth elements are crown glasses, while the second and third are flint glasses, creating a typical telephoto construction with a front positive achromatic group, a rear negative achromatic group, and a field corrector.

S4.2 Additional Results for Baseline Performance

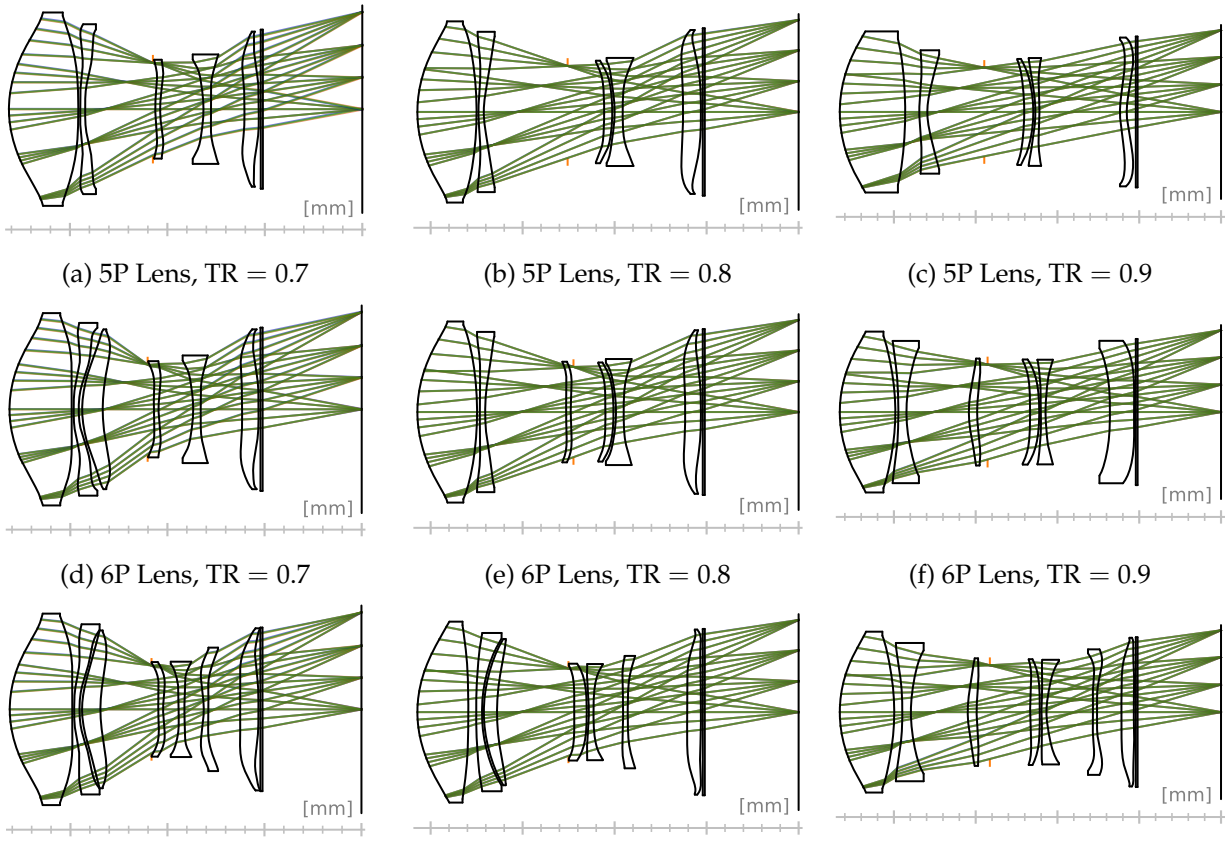
In this section, we compare several lens configurations and evaluate their optical performance in terms of ESR for different TRs.

S4.2.1 Refractive Designs

We evaluate the optical performance of all-refractive designs with 5–7 refractive elements across different TRs (0.7, 0.75, 0.8, 0.85, 0.9). Examples of optimized 2D layouts are shown in Figure S12.

For each configuration and TR, we generate a rough starting point and optimize the lens using the objective in Section S4.1.4, typically for 5000 steps with refractive material optimization followed by another 5000 steps without. We repeat the process when subpar local minima are detected, until all results are consistent with one another. The optical performance of the optimized lenses is reported in terms of ESR in Figure S13, using uniform weights over the FOV and the spectrum in Figure S9. Note that squared ray aberrations are averaged over the fields and wavelengths (weighted), and the ESR is the square root of this average. Thus, ESR correlates strongly with the average spot radius but is not identical.

As shown in Figure S13, the ESR increases monotonically and smoothly with TR. In the high-TR regime, the performance plateaus with an ESR of around 1 µm, which is below the airy disk radius of 1.88 µm for a central wavelength of 550 nm and an f-number of 2.8. Here, chromatic aberrations limit the optical performance, and adding more refractive elements does not provide significant



(a) 5P Lens, TR = 0.7 (b) 5P Lens, TR = 0.8 (c) 5P Lens, TR = 0.9

(d) 6P Lens, TR = 0.7 (e) 6P Lens, TR = 0.8 (f) 6P Lens, TR = 0.9

(g) 7P Lens, TR = 0.7 (h) 7P Lens, TR = 0.8 (i) 7P Lens, TR = 0.9

Figure S12: Examples of 2D layouts for the baseline telephoto lens with 5–7 refractive elements and different TRs. As the TR decreases, we observe the separation of the rear group and, for the 6P and 7P configurations, the formation of a three-element front group.

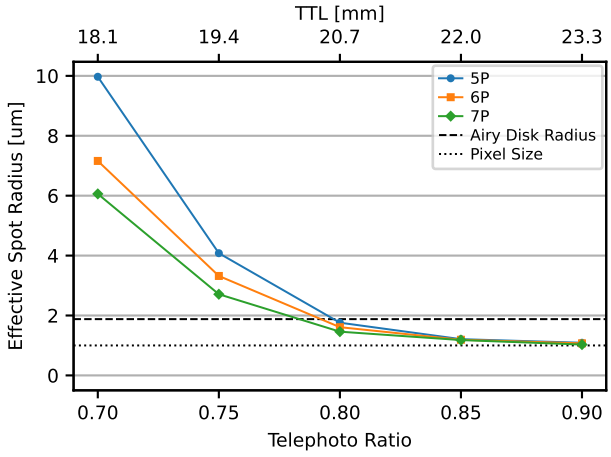


Figure S13: ESR as a function of TR for refractive lenses with 5–7 refractive elements. The labels 5P, 6P, and 7P correspond to the number of refractive elements in the lens. For reference, we include the pixel size of the sensor (1 µm) as well as the airy disk radius at 550 nm (1.88 µm).

improvement.

As the TR decreases, optical performance degrades rapidly due to dominant monochromatic aberrations. In the moderate TR regime (0.75–0.8), adding refractive elements improves performance by 15–20 % per element. The TR required to maintain the same optical performance improves by 0.01–0.02 per additional element. For low TR, as seen in Figure S12, a three-element front group is the key contributor to performance improvement.

S4.2.2 Hybrid Designs

We now explore hybrid designs that combine refractive elements with diffractive surfaces. Diffractive surfaces, assuming all energy goes into the desired diffraction order, have an equivalent Abbe number of -3.45, providing a dispersion opposite and 6–16 times stronger than available moldable refractive materials (see Table S10). However, they can introduce significant scattering losses in broadband illumination, which must be considered in practical implementations. In practice, dual-layered kinoforms can achieve high diffractive efficiency [36].

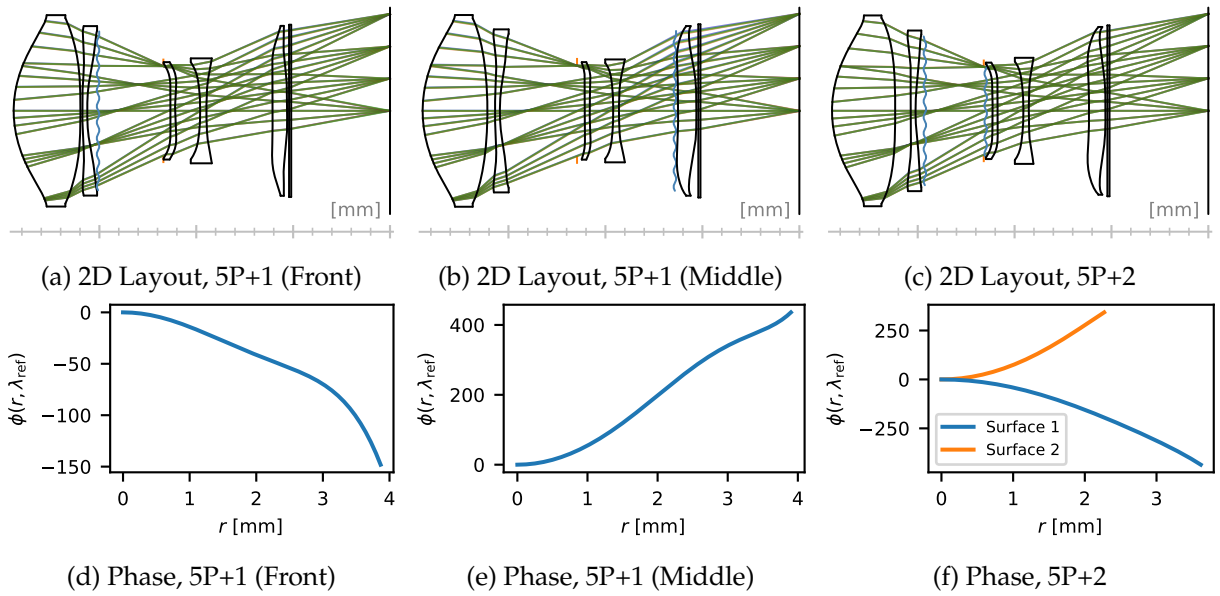
We evaluate three key hybrid design configurations at a fixed TR of 0.75, shown in Figure S14. The first places the diffractive surface at the front, compensating for chromatic aberrations in the front group, making it optimal for high TR. The second places the diffractive surface in the middle, improving performance at low TR by correcting field-dependent monochromatic aberrations, functioning as a flexible field corrector. The third configuration uses two diffractive surfaces, one after the front group and one before the middle, allowing us to assess the impact of multiple diffractive surfaces.

Figure S15 shows the optical performance of the all-refractive 5P and hybrid lens configurations as a function of TR, along with the phase range for each surface. The figure illustrates a "compounding" effect, where adding a second diffractive surface significantly improves performance compared to using just one. This allows for better optimization of surface power, with dispersions compensating for each other, resulting in an increased phase range when two diffractive surfaces are used together.

S4.2.3 Comparison

We compare the ESR of selected configurations in Figure S16, introducing a 6P+1 (front) configuration, where a diffractive surface is placed in front of the lens.

At high TR, refractive-only configurations (5P, 6P, 7P) yield ESR values around 1 µm, close to the pixel size. In contrast, most hybrid configurations with diffractive surfaces perform better, showing an ESR around 0.2 µm and effectively compensating for chromatic aberrations. The exception is the 5P+1 (middle) configuration, which has an ESR of 0.4 µm due to the diffractive surface being too far from the front group. At low TR (0.7), the 5P configuration performs poorly (ESR = 10.0 µm) due to its front group consisting of only two elements. The 5P+2 configuration, with two diffractive surfaces, performs the best (ESR = 4.4 µm), compensating for chromatic dispersion. For intermediate cases, the 5P+1 (front), 6P, and 6P+1 (front) configurations show similar ESR values



(a) 2D Layout, 5P+1 (Front) (b) 2D Layout, 5P+1 (Middle) (c) 2D Layout, 5P+2

(d) Phase, 5P+1 (Front) (e) Phase, 5P+1 (Middle) (f) Phase, 5P+2

Figure S14: Examples of hybrid lenses with multiple diffractive surfaces, all at TR of 0.75. (a, d) In the 5P+1 configuration with the diffractive surface at the front, the surface has a typical negative profile (positive converging lens) to counterbalance the refractive elements' positive dispersion. (b, e) In the 5P+1 configuration with the diffractive surface in the middle, the surface has a positive profile (negative diverging lens). (c, f) In the 5P+2 configuration, the two diffractive surfaces have similar but opposite profiles.

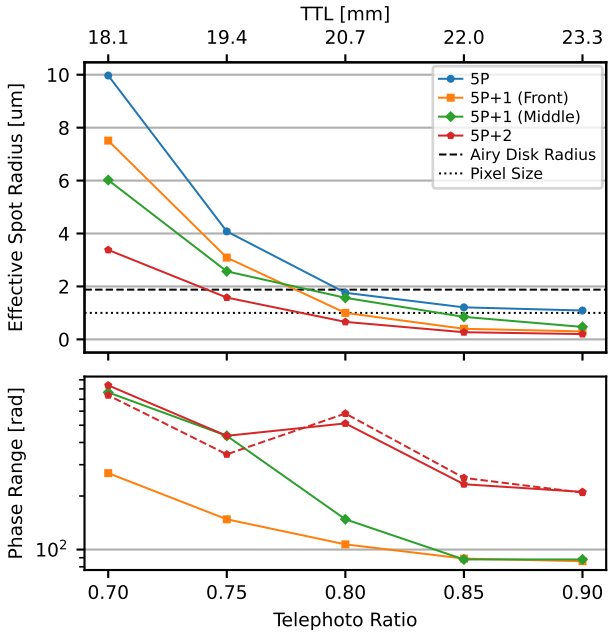


Figure S15: Impact of multiple diffractive surfaces on the optical performance of hybrid lenses, along with their required phase range. In the bottom plot, solid lines represent the phase range of the first diffractive surface, while dashed lines correspond to the second. Only the absolute values of the phase or group delay range are shown; some surfaces exhibit positive power, while others show negative power, as illustrated in Figure S14.

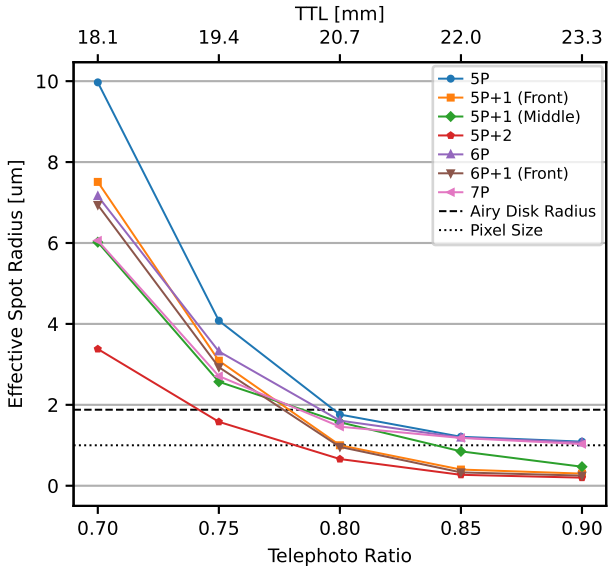


Figure S16: Compilation of optical performance for selected configurations. The ESR is shown as a function of TR for the configurations selected in previous sections.

(7.0–7.5 µm), suggesting that three front elements (one potentially diffractive) are sufficient, with only minor gains from adding a fourth element. Other 6P+1 configurations, where the diffractive element is placed elsewhere, could offer improved performance. Finally, the 5P+1 (middle) and 7P configurations perform similarly (ESR around 6.0 µm), emphasizing the advantage of positioning the diffractive surface correctly.

S4.3 Additional Results for Image Restoration

In this section, we revisit the selection configurations of the previous section and assess their simulated image quality. We investigate image quality pre- and post-image restoration, where a data-driven image restoration model (IRM) is fitted to accommodate each lens.

S4.3.1 Pre-Restoration Image Quality

We propose a methodology to evaluate the raw image quality of a design, specifically focusing on the optical image produced by the system at the sensor. The primary metric is peak signal-to-noise ratio (PSNR), with structural similarity index (SSIM) as a secondary measure. PSNR is preferred for its simplicity, widespread use, and more linear relationship with ESR. While SSIM better reflects perceptual accuracy by considering image structure, the consistent nature of the spatially varying blur in this study makes PSNR a more effective measure.

We test all 100 images from the DIV2K validation set [1], cropping a central 256×256 pixel region from each. Each image is paired with a sensor region of the same size, following the pseudo-random strategy in Section S3.3.1 to ensure equal sensor representation. PSNR and SSIM are averaged across all 100 images. Image quality scores are dataset-dependent and not directly

comparable to optical performance metrics. However, they remain consistent across all optimized designs, enabling comparisons of image quality scores between different configurations and TRs.

In Figure S17, we report the image quality metrics of the raw simulated images for the selected configurations. In Figure S17a, as with the ESR, image quality generally improves with increasing TR. In Figure S17b, image quality improves with decreasing ESR, but the correlation between raw image quality and ESR is not perfect. Some configurations with a larger ESR exhibit better raw image quality than those with a smaller ESR. Notably, all-refractive configurations (5P, 6P, 7P) outperform others for a moderate ESR around 1 µm. This is likely due to the higher chromatic aberrations in these configurations, which significantly degrade the ESR but have a less pronounced impact on the image quality metrics.

In Figure S18, we provide a qualitative comparison of raw simulated captures for the 5P lens. Even the close to diffraction-limited 5P lens with the highest TR of 0.9 does not produce an image free of artifacts compared to the groundtruth image. This can be attributed to the finite size of the point spread functions (PSFs) that is limited by diffraction. We notice a strong degradation in image quality when decreasing the TR from 0.8 to 0.75, and even more so when decreasing it to 0.7 where strong color aberrations become apparent in the high field values. We observe that for a PSNR above around 27 dB, while there may be a noticeable loss of sharpness, the images remain visually similar to the groundtruth. We do not show the other configurations, but note that at equal PSNR values, the images produced by the hybrid lenses tend to be visually similar to the ones from the refractive lenses.

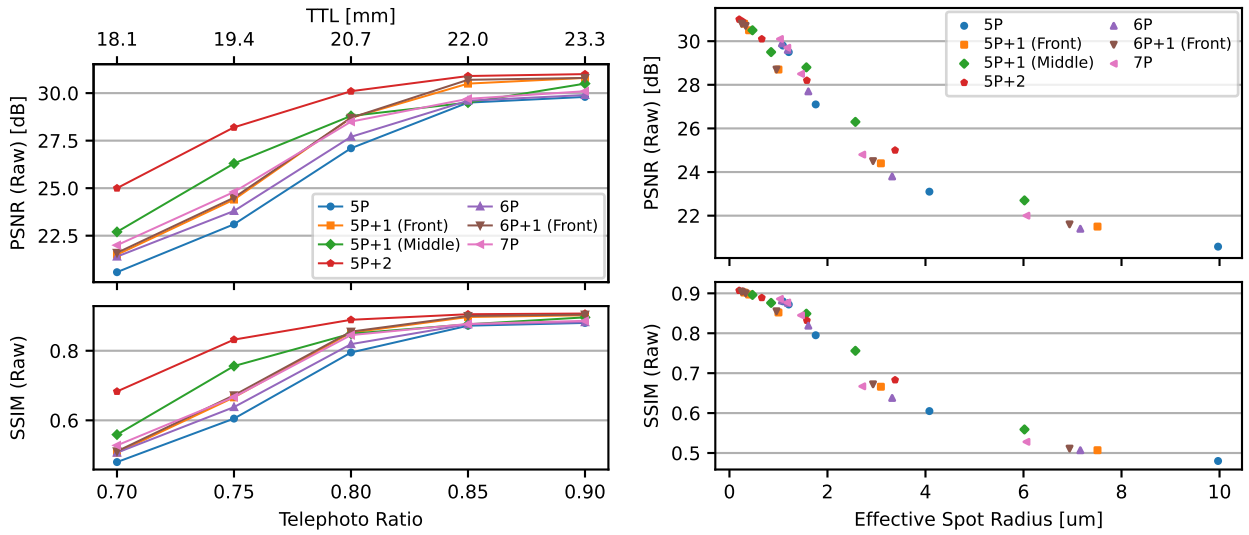
S4.3.2 Post-Restoration Image Quality

Here, we compare the pre- and post-restoration image quality, where a new IRM instance is trained from scratch to adjust to each design. The lenses, optimized for spot size, remain fixed during the training process. We employ the methodology detailed in Section S3.3.1 for training the IRM. We use DIV2K training set [1] (800 images) and apply random crops of 256×256 pixels to each image, and no other data augmentation.

Pre- and post-restoration image quality are compared in Figure S19. Figure S19a shows a significant improvement in PSNR when applying image restoration. To illustrate, the pre-restoration PSNR of the lenses at a TR of 0.9 is surpassed by the post-restoration PSNR of the lenses at a TR of 0.75 for all configurations considered. From Figure S19b, a line corresponding to 1.33 times the pre-restoration PSNR closely fits the post-restoration PSNR. For a pre-restoration PSNR range of 21–30 dB, the post-restoration PSNR is approximately 28–40 dB, a significant improvement of 7–10 dB.

We note that maximizing the restored image quality is not the primary goal of our study, which focuses on the impact of E2E optimization on the optical performance of the imaging system. As an avenue for future work, other IRM architectures or training methodologies could lead to better image quality.

In Figure S20, we provide a qualitative comparison of the restored simulated captures for the



(a) Raw Image Quality vs TR

(b) Raw Image Quality vs Spot Radius

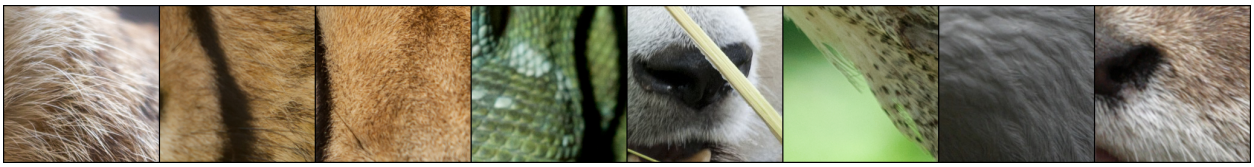
Figure S17: Compilation of raw image quality for selected configurations. (a) The PSNR and SSIM are shown as a function of TR for the configurations selected in previous sections. The trend is generally similar to the one observed for the ESR in Figure S16, with the image quality improving with increasing TR. (b) The PSNR and SSIM are shown as a function of the ESR to illustrate the strong but imperfect correlation between image quality metrics and optical performance metrics.

5P lens. We observe minor blurring starting with the 5P lens with a TR of 0.8, with a PSNR of 36.3 dB and a SSIM of 0.956. Based on these images and more empirical observations, we estimate that a reasonable and conservative threshold for acceptable image quality is a PSNR of 35 dB. This threshold is used to determine the TR at which the image quality is acceptable for all lens configurations.

While IRMs are known to sometimes introduce artifacts such as ringing, color artifacts, or over-sharpening, we do not observe any such artifacts in the restored images.

S4.4 Additional Results for End-to-End Optimization

In this section, we provide additional results for the E2E-optimized scenario, where smartphone telephoto lenses are optimized alongside the IRM to maximize the restored image quality. We evaluate the impact of E2E optimization on the image quality of the selected configurations, comparing the optical response of the optimized systems to that of the spot-optimized designs. Additionally, we use an independent validation to assess the image quality of the optimized designs, verifying whether the improvement in image quality is consistent across different simulation methods.

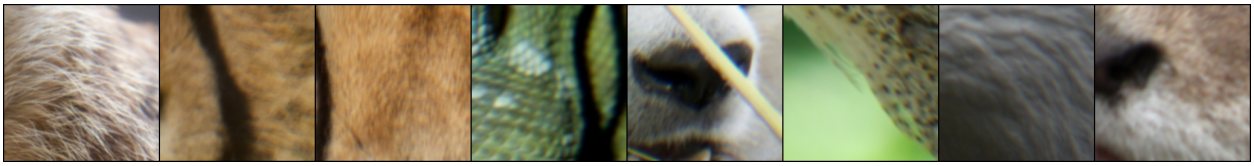


(a) Groundtruth

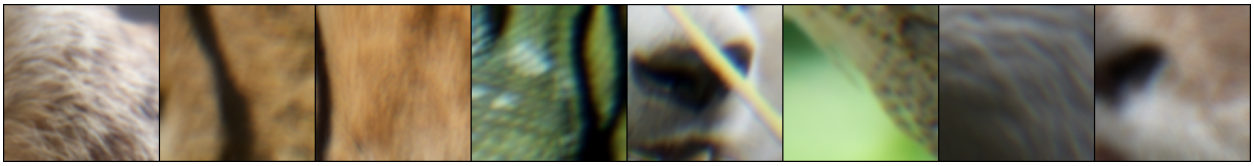
(b) 5P lens, TR $= 0.9$, PSNR $= 29.8$ dB, SSIM $= 0.880$

(c) 5P lens, TR $= 0.85$, PSNR $= 29.5$ dB, SSIM $= 0.872$

(d) 5P lens, TR $= 0.8$, PSNR $= 27.1$ dB, SSIM $= 0.795$

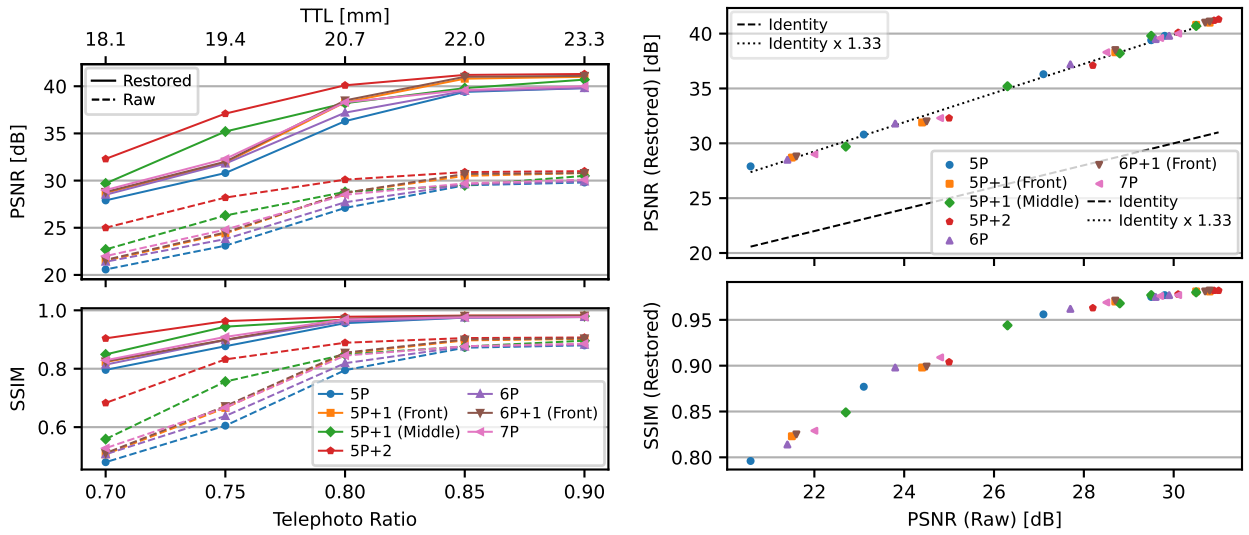


(e) 5P lens, TR $= 0.75$, PSNR $= 23.1$ dB, SSIM $= 0.605$



(f) 5P lens, TR $= 0.7$, PSNR $= 20.6$ dB, SSIM $= 0.480$

Figure S18: Qualitative comparison of raw simulated captures for the 5P configuration. Going from the leftmost to the rightmost image, the simulated captures correspond to progressively higher field values, with the first and last images closest to the center and corners of the sensor, respectively. Ground truth image crops are shown in (a), while the simulated captures produced by the 5P lens with different TRs are shown in (b)–(f). The averaged PSNR and SSIM scores are reported for each configuration.



(a) Restored Image Quality vs TR

(b) Restored Image Quality vs Raw PSNR

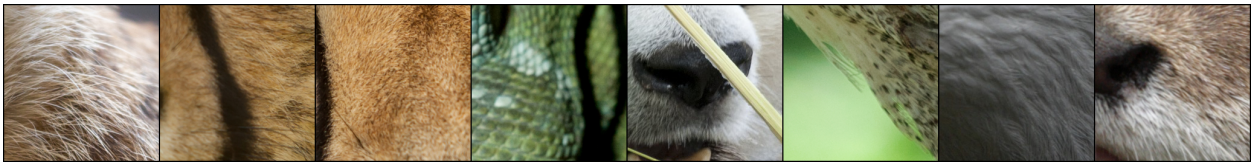
Figure S19: Raw and restored image quality for selected configurations. (a) The PSNR and SSIM are shown as a function of TR for the configurations selected in previous sections. (b) The PSNR and SSIM are shown as a function of the raw PSNR. For reference, the dashed line in the PSNR plot corresponds to the identity line. The dotted line, set to 1.33 times the identity line, corresponds to an approximate fit to the data.

S4.4.1 Impact of End-to-End Optimization on Image Quality

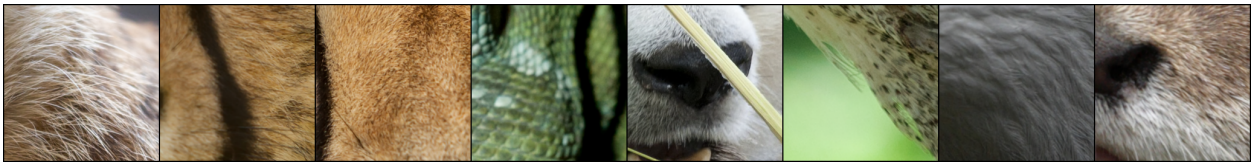
In Figure S21, we report the image quality of the E2E-optimized lenses and compare it to the spot-optimized designs. In the E2E-optimized systems, by design and as can be verified in Figure S21b, the ESR is always larger than in the spot-optimized designs. However, the restored PSNR is improved in most cases, and the improvement is substantial in many of them. The improvement in PSNR is commonly above 2 dB and can be as high as 4 dB. We note that the improvement in PSNR is generally largest in the low- and moderate-TR regimes, where the optical performance is dominated by geometric aberrations. In rare cases with very large spot radius, the change in PSNR is negative, denoting a failure of the E2E optimization process. This happens when the PSFs are so large that a significant proportion of their energy hits outside of the PSF array of 49×49 µm, which derails the optimization process as it is not differentiable nor accounted for.

Figure S22 shows the restored image quality metrics as a function of the ESR of the spot-optimized lenses. The ESR of the spot-optimized lenses provides a measure of the degree of geometric aberrations in a lens, as well as the potential of the lens configuration in correcting these aberrations based on its degrees of freedom. The figure shows that the improvement in image quality from E2E optimization is generally at its highest when geometric aberrations are significant, around the Airy disk radius or more. Some configurations are better at maximizing restored PSNR than others for a given ESR, like the 5P+1 (middle) configuration compared to the 7P one, suggesting that the degrees of freedom are better suited to achieve particular PSFs.

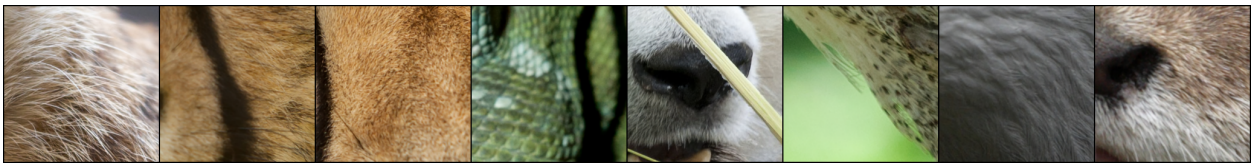
In Figure S23, we provide a qualitative comparison of the restored simulated captures for the



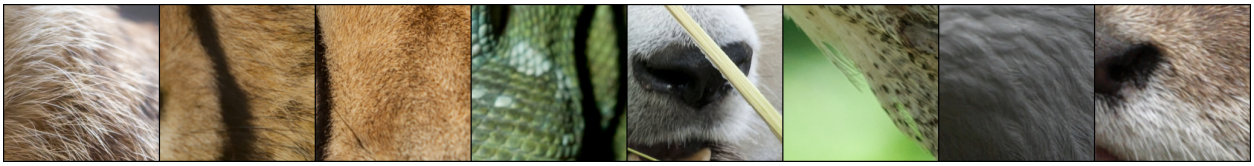
(a) Groundtruth



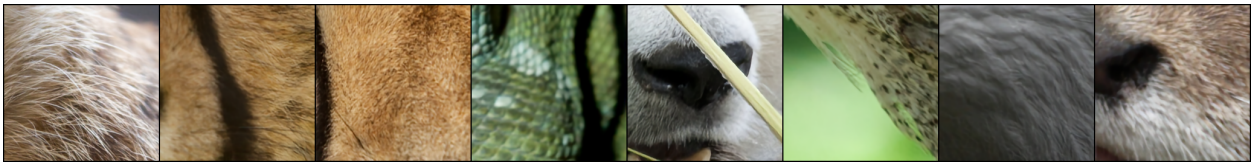
(b) 5P lens, TR $= 0.9$, PSNR $= 39.8$ dB, SSIM $= 0.977$



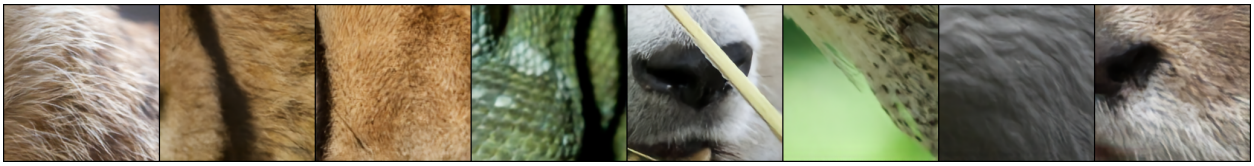
(c) 5P lens, TR $= 0.85$, PSNR $= 39.4$ dB, SSIM $= 0.975$



(d) 5P lens, TR $= 0.8$, PSNR $= 36.3$ dB, SSIM $= 0.956$

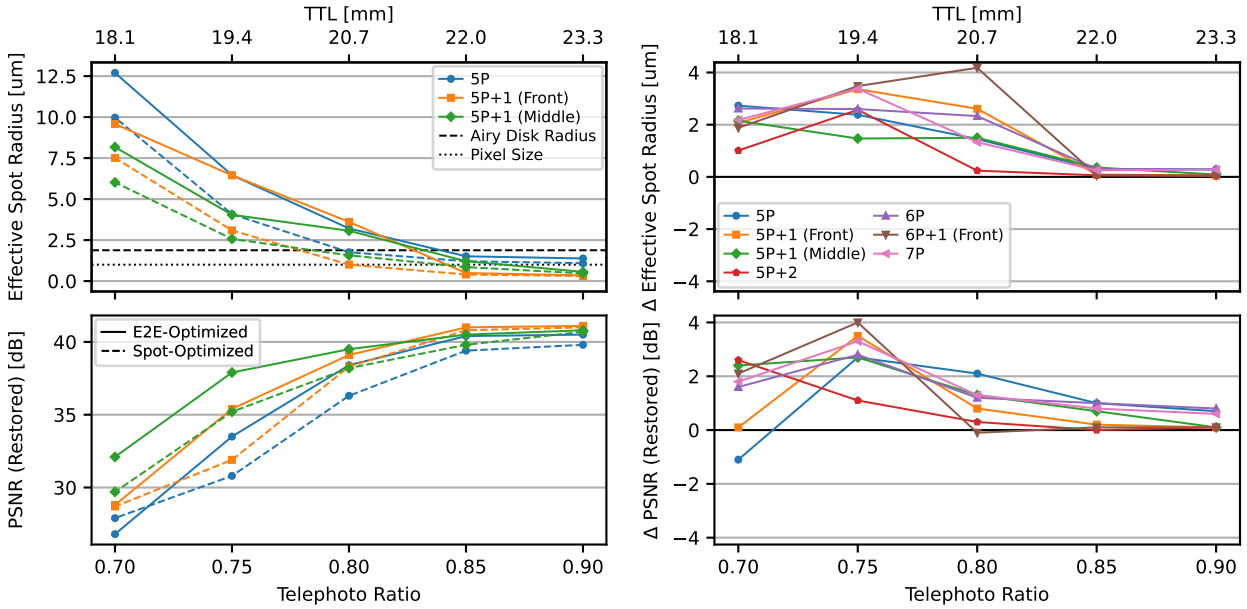


(e) 5P lens, TR $= 0.75$, PSNR $= 30.8$ dB, SSIM $= 0.877$



(f) 5P lens, TR $= 0.7$, PSNR $= 27.9$ dB, SSIM $= 0.796$

Figure S20: Qualitative comparison of restored simulated captures for the 5P lens. See Figure S18 for a description of the images. When used alongside image restoration, 5P lenses with a TR of 0.9 and 0.85 produce images that are nearly indistinguishable from the groundtruth images. Minor but noticeable blurring is present for a TR of 0.8. Below this value, the images are clearly degraded. We note that the restored images do not present visible artifacts other than blurring.



(a) Spot and PSNR of E2E-Optimized Lenses

(b) Impact of E2E Optimization

Figure S21: Image quality of the E2E-optimized designs compared to their spot-optimized counterparts. (a) The ESR and PSNR are given as a function of TR for a subset of the configurations. (b) The difference in ESR and PSNR between the E2E- and spot-optimized lenses is given for all configurations of interest.

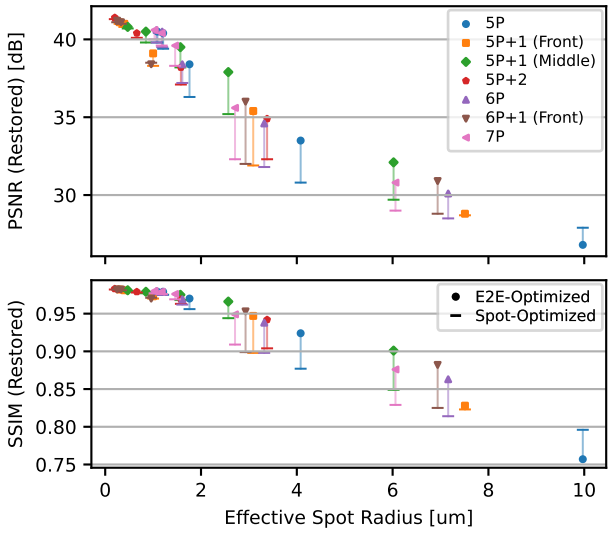
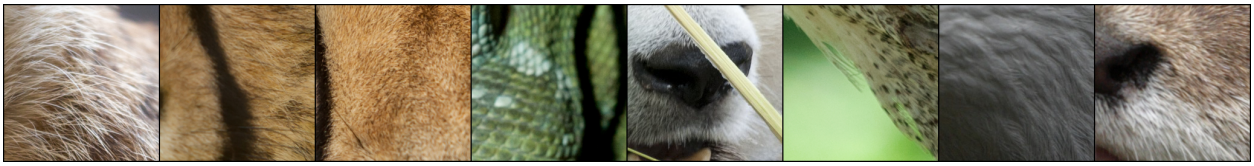
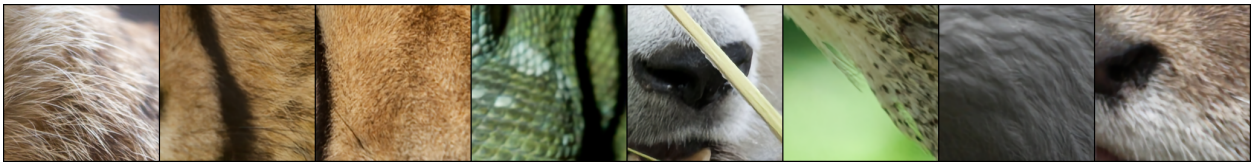


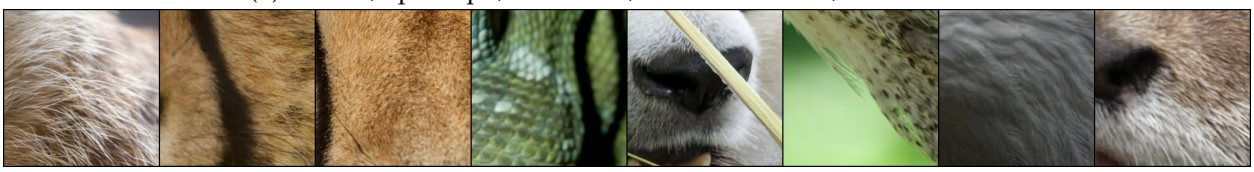
Figure S22: The image quality metrics are plotted against the ESR of the spot-optimized lenses. This shows how the improvement in image quality from E2E optimization varies according to the strength of the geometric aberrations.



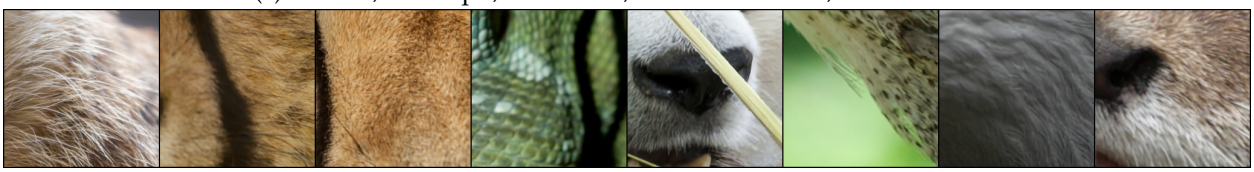
(a) Groundtruth



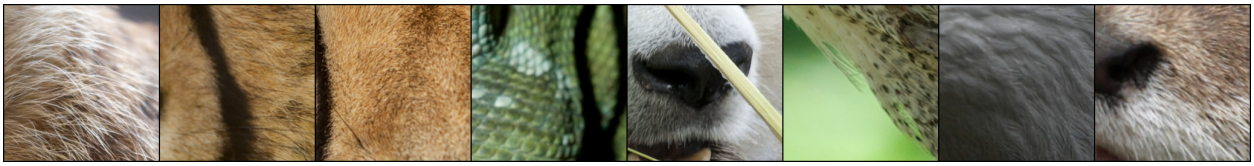
(b) 5P lens, Spot-Opt., TR = 0.75, PSNR = 30.8 dB, SSIM = 0.877



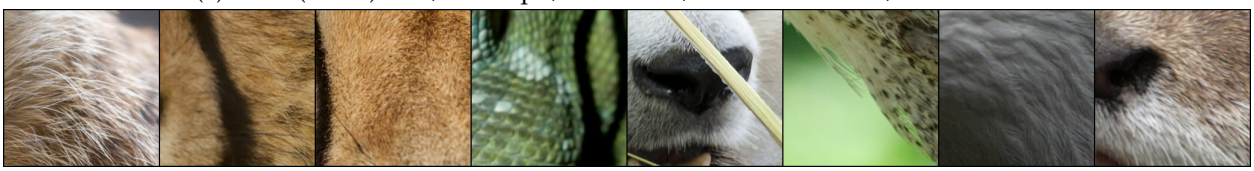
(c) 5P lens, E2E-Opt., TR = 0.75, PSNR = 33.5 dB, SSIM = 0.924



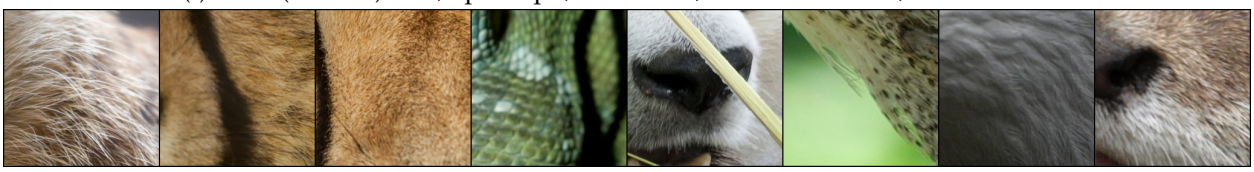
(d) 5P+1 (Front) lens, Spot-Opt., TR = 0.75, PSNR = 31.9 dB, SSIM = 0.898



(e) 5P+1 (Front) lens, E2E-Opt., TR = 0.75, PSNR = 35.4 dB, SSIM = 0.947



(f) 5P+1 (Middle) lens, Spot-Opt., TR = 0.75, PSNR = 35.2 dB, SSIM = 0.944



(g) 5P+1 (Middle) lens, E2E-Opt., TR = 0.75, PSNR = 37.9 dB, SSIM = 0.966

Figure S23: Qualitative comparison of restored simulated captures for the 5P and 5P+1 (front and middle) lenses, either spot-optimized or E2E-optimized, shown for a TR of 0.75. See Figure S18 for a description of the images. The restored images produced by the E2E-optimized lenses are generally sharper. This is to the exception of the rightmost images—corresponding to corners of the image sensor—as the E2E-optimized lenses tend to sacrifice sharpness near the edge of the FOV.

5P and 5P+1 (front and middle) lenses, either spot-optimized or E2E-optimized, shown for a TR of 0.75.

S4.4.2 Optimized Optical Response in End-to-End Systems

In this section, we are interested in understanding the optimized optical response of the E2E systems and comparing it to the optical response of the spot-optimized designs.

In Figure S24, we report the difference in raw PSNR between the E2E- and spot-optimized lenses. We observe that the raw PSNR tends to improve for the E2E-optimized lenses, but not always. The improvement is generally up to 5 %, equivalent to 1.3 dB around a raw PSNR of 26 dB. Therefore, the improvement in raw PSNR partly explains the improvement in restored PSNR, but not completely as the improvement in restored PSNR is generally larger, in the range of 12 % in the peak region. Additionally, many designs with a negligible improvement in raw PSNR still show a substantial improvement in restored PSNR, indicating that the E2E optimization process is capable of improving the restored image quality beyond the raw image quality.

Next, we qualitatively compare a subset of E2E-optimized designs with their spot-optimized counterparts. In Figure S25, we present RGB PSFs for the 5P, 5P+1 (front), and 5P+1 (middle) configurations for both types of designs. Two key observations emerge. First, E2E-optimized designs have narrower central peaks but larger support and longer tails, spreading the energy more widely. This explains the difference in ESR, as energy in the tails disproportionately impacts ESR due to the squared distance weighting. Second, E2E-optimized designs prioritize central and intermediate image quality at the expense of the corners, where image content is less crucial, reflecting a common trade-off in image restoration aimed at maximizing overall image quality across the sensor.

Figure S26 shows the 2D layouts for the same lenses as in Figure S25, where it can be seen that the overall lens structure is preserved between spot- and E2E-optimized designs. Some changes are consistent across all configurations and TRs, such as the second refractive element having a more pronounced curvature in the E2E-optimized designs and the field corrector element having a flatter shape.

S4.4.3 Validation of Image Quality

In this section, we use the commercial optical design software Synopsys CODE V [32] to validate the image quality metrics obtained using our framework.

As noted in Section S1.3.4, our ray-tracing operations closely match those in CODE V when the aperture stop is in object space and the lens is free of vignetting. However, we expect less agreement here since the aperture stop is inside the lens, requiring ray aiming, and vignetting is present. Most importantly, our framework uses geometric PSFs with a heuristic for diffraction, while CODE V performs full diffraction calculations. Thus, we aim to quantify the differences in image quality metrics between the two frameworks and confirm that E2E optimization improvements are consistent.

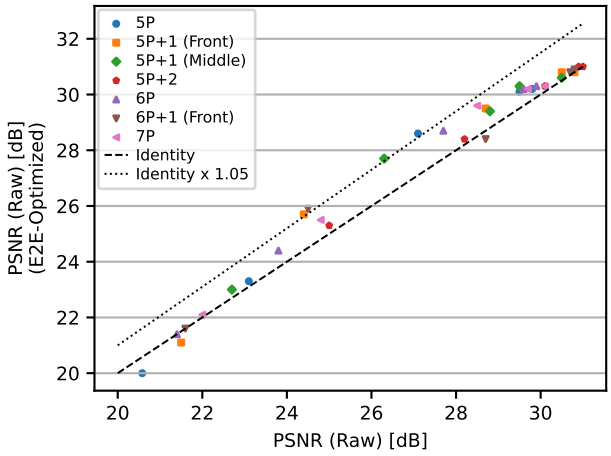
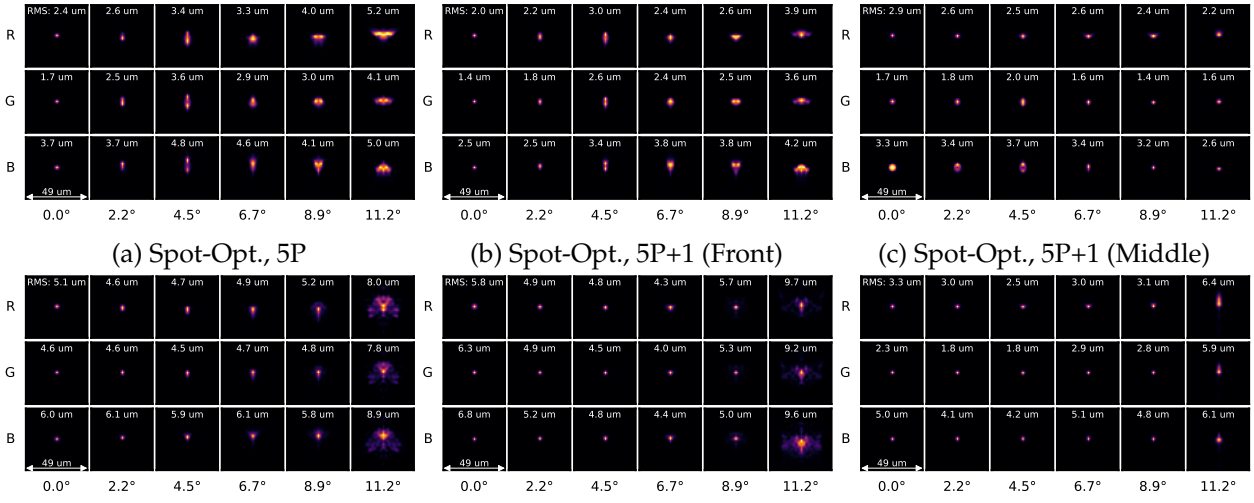


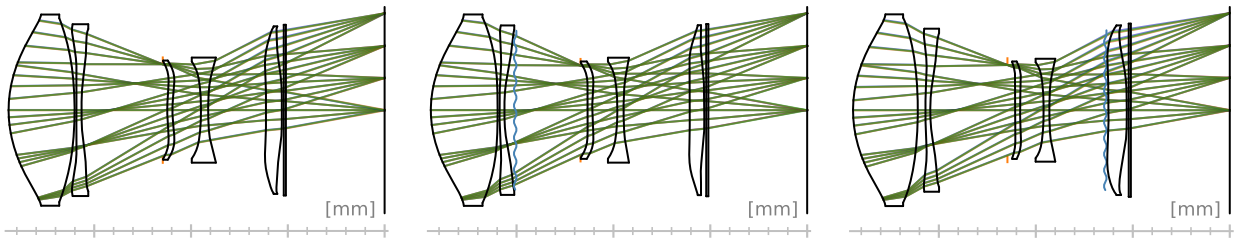
Figure S24: Pre-restoration PSNR of E2E-optimized lenses compared to the spot-optimized lenses. The raw PSNR tends to improve for the E2E-optimized lenses, though not in all cases, by up to 5 %.



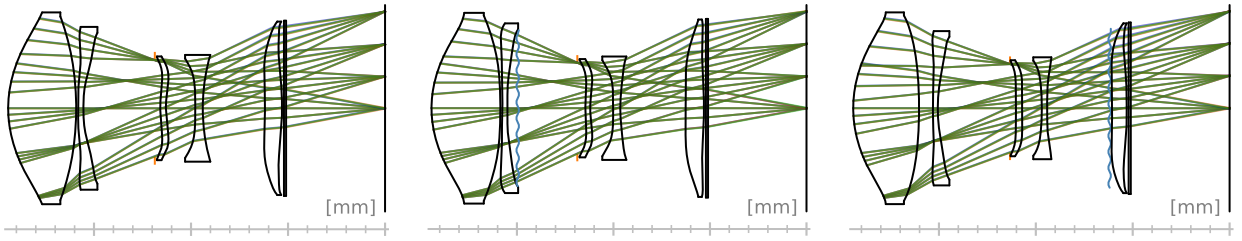
(a) Spot-Opt., 5P (b) Spot-Opt., 5P+1 (Front) (c) Spot-Opt., 5P+1 (Middle)

(d) E2E-Opt., 5P (e) E2E-Opt., 5P+1 (Front) (f) E2E-Opt., 5P+1 (Middle)

Figure S25: Examples of PSFs from the E2E optimization process compared to their spot-optimized counterparts for a TR of 0.75. We note that the E2E-optimized designs have a narrower central peak than the spot-optimized designs, with the trade-off being a larger support and longer tails. Additionally, the E2E-optimized designs sacrifice image quality in the corners of the sensor for better image quality in intermediate regions. These trends are consistent across all configurations and TRs.



(a) Spot-Opt., 5P (b) Spot-Opt., 5P+1 (Front) (c) Spot-Opt., 5P+1 (Middle)



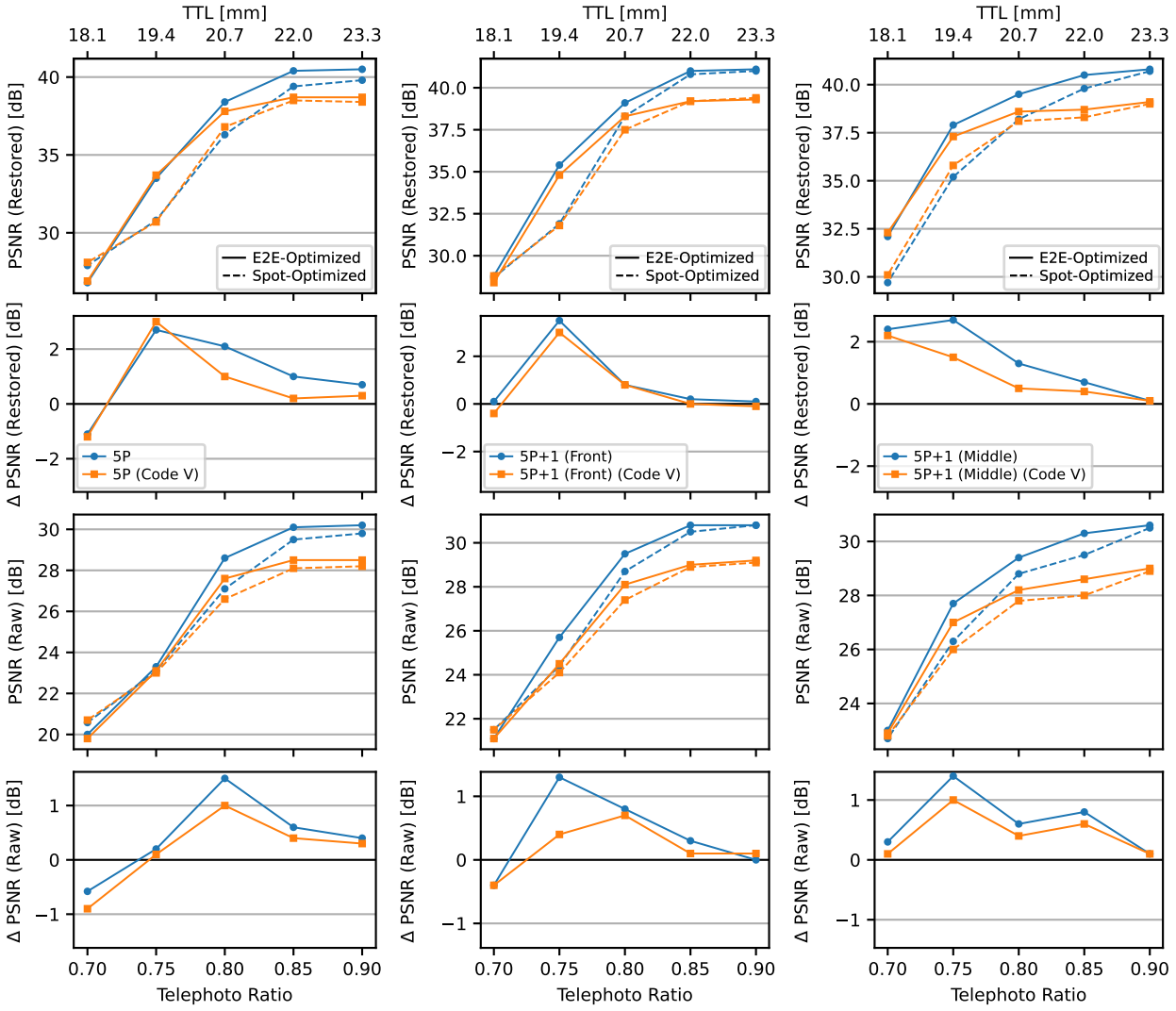
(d) E2E-Opt., 5P (e) E2E-Opt., 5P+1 (Front) (f) E2E-Opt., 5P+1 (Middle)

Figure S26: Examples of layouts from the E2E optimization process, compared to their spot-optimized counterparts, for a TR of 0.75. While the overall structure of the lens is preserved between the spot-optimized and E2E-optimized designs, some of the individual elements change noticeably in shape.

We first export our spot- and E2E-optimized designs to CODE V, along with dispersion data for each refractive material to ensure consistent refractive index calculations, as detailed in Section S1.3.4. From CODE V, we extract the diffraction PSFs for 11 wavelengths and 11 equidistant field values to fit our framework's format. The extracted CODE V PSFs are postprocessed by calculating the wavelength-averaged centroid, recentering each PSF relative to its field centroid, rescaling to a 1 µm pixel pitch, cropping to a 49×49 µm array, and normalizing by area. This aligns the CODE V PSFs with our format. We then re-train the IRM from scratch with the CODE V PSFs, using the same methodology as in Section S3.3.1, and evaluate image quality metrics on the DIV2K validation set.

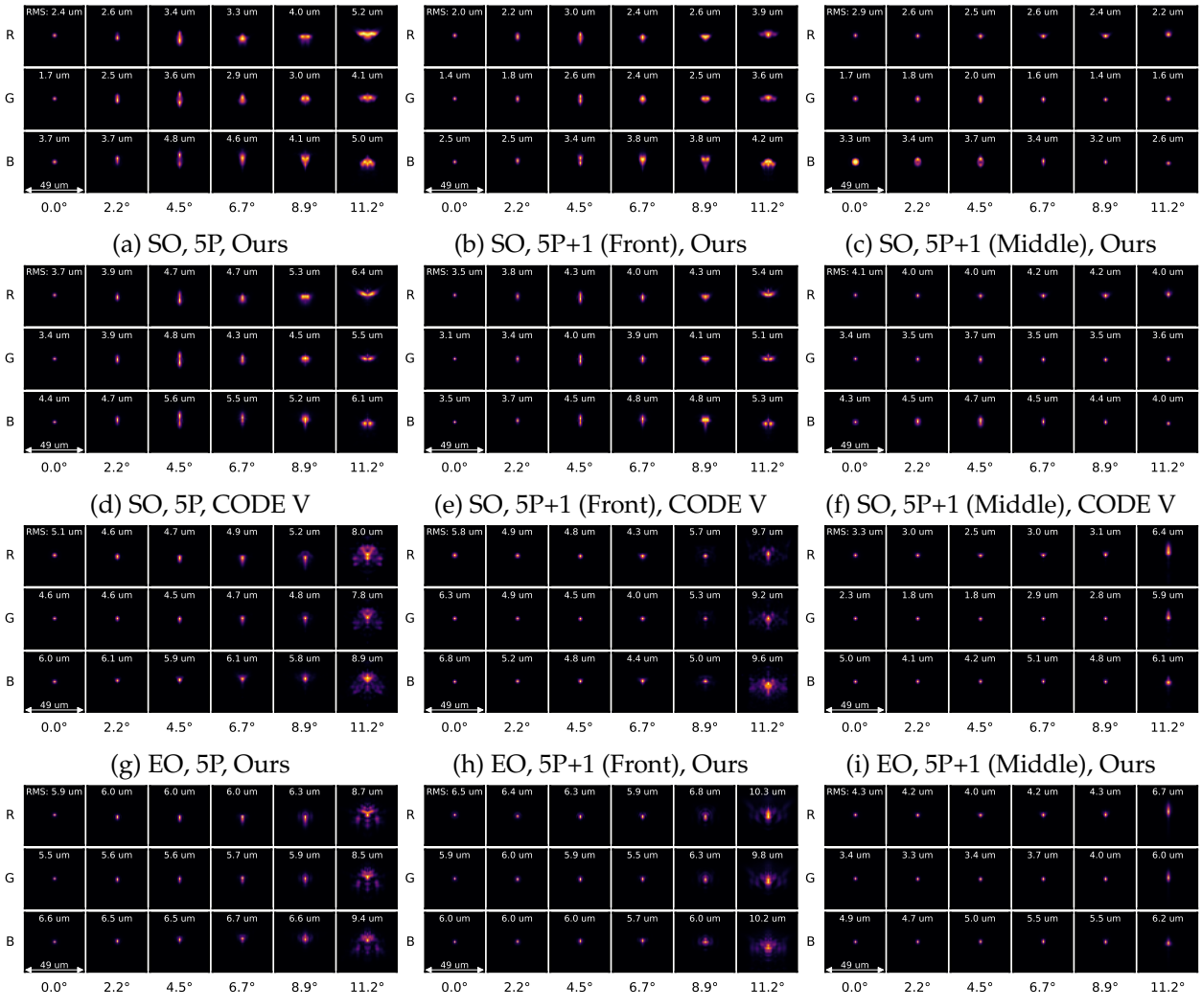
Figure S27 compares the image quality metrics with PSFs from our framework and CODE V for the 5P, 5P+1 (front), and 5P+1 (middle) configurations. There is a good agreement between the image quality metrics obtained using our framework and CODE V, although the metrics are generally lower when using CODE V as the lenses approach diffraction-limited performance (around a raw PSNR of 26 dB). Most importantly, the improvements from E2E optimization on post-restoration image quality are mostly preserved when using CODE V, although in some cases the improvement is decreased by up to 1 dB.

Figure S28 compares the RGB PSFs for our framework and CODE V for the 5P, 5P+1 (front), and 5P+1 (middle) configurations, for a TR of 0.75. The PSFs are similar in terms of the general shape and size. The differences are mostly in low-frequency details, which can be mainly attributed to CODE V's full diffraction calculation.



(a) 5P Lens (b) 5P+1 (Front) Lens (c) 5P+1 (Middle) Lens

Figure S27: Validation of image quality metrics and improvement from E2E optimization for the 5P, 5P+1 (front), and 5P+1 (middle) configurations. We export all lens designs to Synopsys CODE V, extract the diffraction PSFs, rerun the training of the IRM using these PSFs, and evaluate the image quality metrics on the validation set.



(a) SO, 5P, Ours (b) SO, 5P+1 (Front), Ours (c) SO, 5P+1 (Middle), Ours

(d) SO, 5P, CODE V (e) SO, 5P+1 (Front), CODE V (f) SO, 5P+1 (Middle), CODE V

(g) EO, 5P, Ours (h) EO, 5P+1 (Front), Ours (i) EO, 5P+1 (Middle), Ours

(j) EO, 5P, CODE V (k) EO, 5P+1 (Front), CODE V (l) EO, 5P+1 (Middle), CODE V

Figure S28: Comparison of PSFs obtained using our framework and CODE V for the 5P, 5P+1 (front), and 5P+1 (middle) configurations. The PSFs are shown for the spot-optimized (SO) and E2E-optimized (EO) designs.

S5 Additional Results for Microscope Objective Lenses

In this section, we provide additional information on the study on the task-driven design of microscope objective lenses (MOLs).

In Section S5.1, we detail the problem formulation and baseline design for microscope objective lenses. In Section S5.2, we present additional results and analysis.

S5.1 Problem Formulation and Baseline Design

This study focuses on MOL with moderate numerical aperture (NA), a complex lens design problem with challenges distinct from those faced by smartphone telephoto lenses. MOLs typically feature a small field of view (FOV) but a proportionally large aperture, and are constructed from multiple lens elements, often spherical and made from conventional glass materials. As such, MOL have access to a broader range of optical materials, placing greater emphasis on color correction to accommodate a wider spectrum of light. Similar to how the telephoto ratio (TR) was varied in the telephoto lens study, here we vary the working distance (WD), where longer WDs enable more applications but significantly change the design structure. As discussed in [5], the optimal structure of the lens changes with the WD, with Lister or Double-Gauss designs being preferred for shorter WDs and retrofocus designs for longer WDs. This enables us to explore a wider range of design structures to draw meaningful conclusions.

We detail the specifications in Section S5.1.1, the spectrum in Section S5.1.2, the constraints and optimization objective in Section S5.1.3, and present a baseline design in Section S5.1.4.

S5.1.1 Specifications

For the purpose of this study, we consider the specifications listed in Table S12.

We base our design specifications on an NA of 0.4, a magnification of 20× and an Atlas 16.8MP camera. Based on a tube lens with a focal length of 200 mm, the effective focal length (EFL) is set to 10 mm to result in a magnification of 20×. The pixel pitch is set to 6.9 µm after considering 2x2 binning, which amounts to an equivalent size of 0.345 µm in object space, well below the Airy disk radius of 0.839 µm for the given NA of 0.4 and wavelength of 550 nm. Using patches of 256×256 px for the image simulation, this results in 768 (10×6) discrete possibilities for selecting a sensor region, assuming no overlap between sensor tiles as in Section S3.3.1. We use a curvature solve on the last surface to enforce the desired focal length.

S5.1.2 Visible Spectrum

In Figure S29, we illustrate the quantum efficiency curves of the Atlas 16.8MP camera, which we use to optimize and evaluate the MOL. The spectrum corresponds to a regular grid of 11 wavelengths between 400–650 nm. The average wavelength is 536.6 nm, and the standard deviation is 76.3 nm,

Table S12: Summary of specifications for the MOL study.

Parameter	Value
Sensor specifications	
Sensor	4/3″ Sony IMX387
Sensor resolution	2560 × 1536 px (down from original resolution)
Pixel pitch	6.9 µm (after considering 2 × 2 binning)
Sensor diagonal	20.6 mm
First-order specifications	
Field of view	5.90°
Nominal image height	10.3 mm
Nominal effective focal length (EFL)	10 mm
Nominal f-number	1.25
Entrance pupil diameter	8 mm
Nominal spectrum	Detailed in Section S5.1.2
Image quality specifications	
Distortion	$\leq 2\,\%$ across the field
Manufacturability specifications	
Element thickness	≥ 1 mm
Thickness to center thickness ratio	$1/3 \leq$ ratio ≤ 3 across element diameter
Air gap spacing	≥ 0.5 mm
Surface normals of refractive elements	$\leq 30°$
Angles of incidence/refraction	$\leq 60°$
Working distance	Variable (4–20 mm)
Miscellaneous specifications	
Refractive materials	104 recommended glasses from Ohara catalog

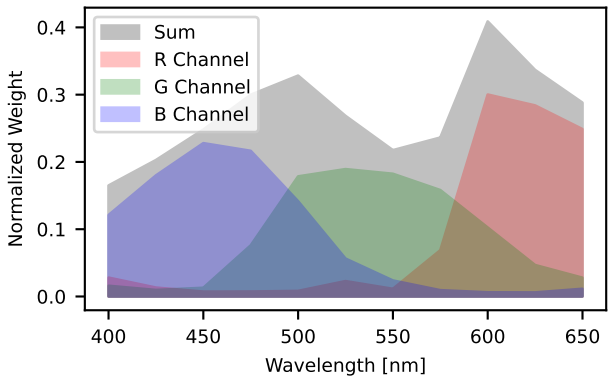


Figure S29: We use the quantum efficiency curves of the Atlas 16.8MP camera to define the spectrum for MOLs. We consider a regular grid of 11 wavelengths between 400–650 nm. The monochromatic version, obtained by summing the spectra for the three RGB channels, is used when optimizing the lenses for spot size.

1.71× the one reported for the telephoto lens study (Section S4.1.2). In Table S13, we list the weights for the RGB channels.

Table S13: Normalized weights for the RGB channels of the MOL. This corresponds to the weight matrix $\mathbf{W}_w$ in Section S1.5.1 that is used to combine the PSFs from individual wavelengths into RGB PSFs.

Wavelength (nm)	R channel	G channel	B channel	Average
400	0.0280	0.0159	0.1210	0.0550
425	0.0133	0.0097	0.1799	0.0676
450	0.0072	0.0129	0.2279	0.0827
475	0.0074	0.0760	0.2165	0.1000
500	0.0083	0.1785	0.1416	0.1094
525	0.0228	0.1893	0.0563	0.0895
550	0.0113	0.1824	0.0238	0.0725
575	0.0688	0.1582	0.0093	0.0788
600	0.3002	0.1025	0.0062	0.1363
625	0.2840	0.0467	0.0060	0.1122
650	0.2488	0.0279	0.0114	0.0960

S5.1.3 Constraints and Optimization Objective

Based on the specifications listed in Table S12, we establish the constraints and optimization objective for the MOL problem. A summary of the constraints and optimization objective is provided in Table S14.

As in the telephoto lens study, we use the same optimization objective in all experiments, to the sole exception of interchanging ℓ_{TRA} and ℓ_{GTRA} for the spot- and mean squared error (MSE)-optimized scenarios, respectively.

S5.1.4 Baseline Design

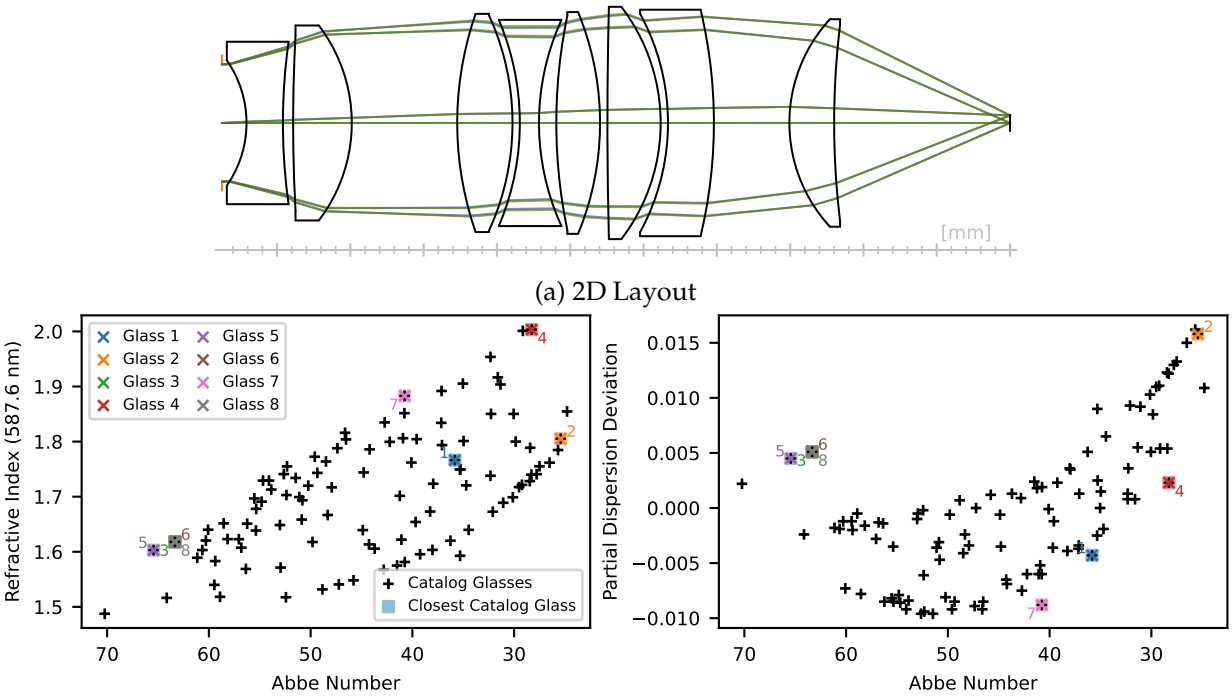
In this section, we present a baseline design for the telephoto lens and detail our design choices, which we represent in Figure S30. This baseline design was optimized with the optimization objective detailed in Section S5.1.3 for spot size, and corresponds to a WD of 12 mm. The lens is composed of eight spherical refractive elements, and is representative of all the lenses that are presented in this study. As depicted in Figure S30b, we consider 104 recommended glasses from the Ohara catalog. Unlike the telephoto lens study, we do not make the assumption that the available materials have normal partial dispersion.

Aperture stop We locate the aperture stop in front of the first refractive element, such that no vignetting is present and iterative ray aiming is not required.

Solves We employ a single solve on the last optimizable surface curvature to enforce the desired focal length. The last spacing is fixed to the desired WD. Note that our definition of the WD is the distance between the vertex of the last surface and the sensor; the actual distance between the two

Table S14: Summary of constraints and optimization objective for the MOL problem. In our experiments, we note that either ℓ_{TRA} or ℓ_{GTRA} is used at a given time, whether the lens is spot-optimized or MSE-optimized. All other residuals are used in all experiments with the same weights except where mentioned.

Residuals	Description	Notes
ℓ_{TRA}	Transverse ray aberrations residuals. See Section S2.2.1.	Only used in the spot-optimized scenario.
ℓ_{GTRA}	Generalized transverse ray aberrations residuals. See Section S3.1.	Used in the mean squared error (MSE)-optimized scenario.
ℓ_{RP}	Ray path residuals. See Section S2.2.2.	In refractive elements, the lower threshold is set to the highest value between 1 mm or CT/3, where CT is the center thickness. The upper threshold is set to the lowest value between 4 mm or 3CT. In air gaps, a lower threshold is set to 0.5 mm.
ℓ_{RA}	Ray angle residuals. See Section S2.2.2.	The threshold is set to $\pm 60°$.
ℓ_{SN}	Surface normal residuals. See Section S2.2.2.	The threshold is set to $30°$.
ℓ_{D}	Distortion residuals. See Section S2.2.3.	The threshold is set to $\pm 2\,\%$.
ℓ_{GMD}	Glass mesh distance residuals. See Section S2.2.4.	The maximum edge threshold is set to 1.25.
ℓ_{GCD}	Glass catalog distance residuals. See Section S2.2.4.	



(a) 2D Layout

(b) Refractive Materials

Figure S30: Baseline MOL design for 8P and WD of 12 mm. (a) 2D layout of the MOL. The lens is composed of eight spherical refractive elements. The aperture stop is located in front of the first refractive element and there is no vignetting. (b) Optimized materials.

components is generally slightly smaller when the last surface is concave. The two corresponding variables are removed from the optimization problem.

Optimizable variables In total, the lens has 49 optimizable variables when the materials are optimized and 31 when the materials are fixed. This corresponds to 16 spacings (out of 17), 15 curvatures (out of 16), and 6×3 glass variables.

S5.2 Additional Results

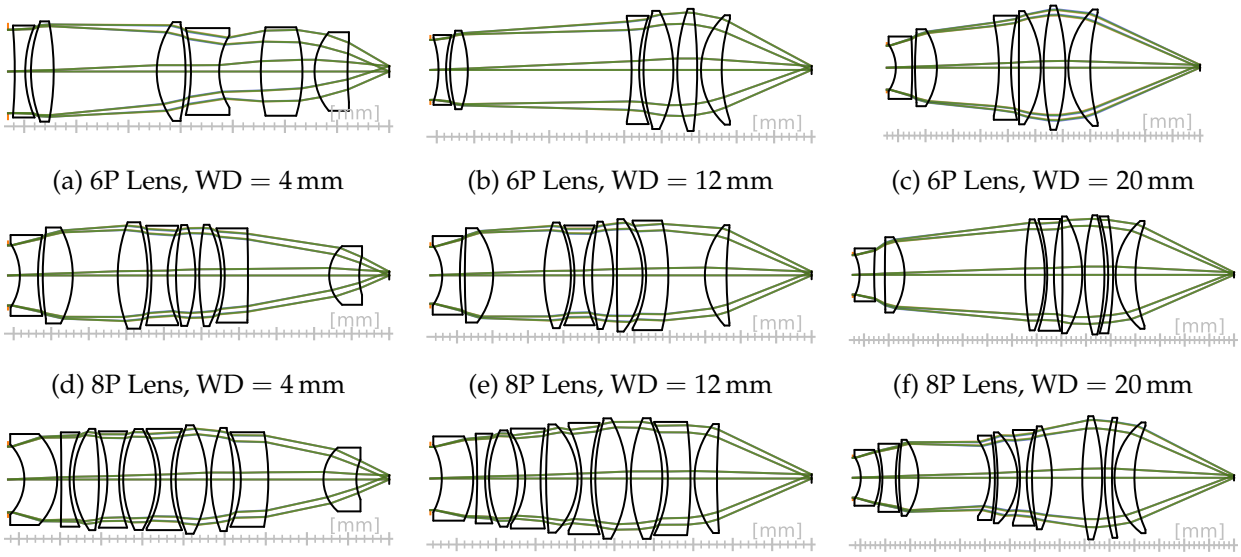
In this section, we present additional results and analysis for the MOL study.

In Figure S31, we present examples of 2D layouts for MOL with 6–10 refractive elements and different WDs. At a WD of 12 mm, we observe a retrofocus structure for the designs, which becomes more pronounced at a WD of 20 mm.

In Figure S32, we compare the effective spot radius (ESR) for the spot-optimized scenario for all designs. The ESR monotonically and smoothly increases with the WD, forming a coherent picture and suggesting that all designs have a performance close to the optimum. The 6P lenses perform much worse than the 8P and 10P lenses, indicating that they lack the necessary degrees of freedom to achieve a good performance. The 8P and 10P lenses have similar performance, with the 10P lenses performing slightly better.

In Figure S33a, we compare the ESR between the MSE- and spot-optimized scenarios. The ESR in the MSE-optimized scenario is always larger than in the spot-optimized scenario, as expected. In Figure S33b, we compare the raw peak signal-to-noise ratio (PSNR) between the MSE- and spot-optimized scenarios. The improvement in the raw PSNR is up to 9 %.

In Figure S34, we present examples of point spread functions (PSFs) for the MSE- and spot-optimized scenarios. Despite the designs being optimized for raw image quality rather than restored image quality, we observe similar trends as in the telephoto lens study: the MSE-optimized designs have a narrower central peak than the spot-optimized designs, with the trade-off being a larger support and longer tails. Additionally, the MSE-optimized designs sacrifice image quality in the corners of the sensor for better image quality in intermediate regions. These trends are consistent across all configurations and WDs.



(a) 6P Lens, WD = 4 mm (b) 6P Lens, WD = 12 mm (c) 6P Lens, WD = 20 mm

(d) 8P Lens, WD = 4 mm (e) 8P Lens, WD = 12 mm (f) 8P Lens, WD = 20 mm

(g) 10P Lens, WD = 4 mm (h) 10P Lens, WD = 12 mm (i) 10P Lens, WD = 20 mm

Figure S31: Examples of 2D layouts for MOL with 6–10 refractive elements and different WDs.

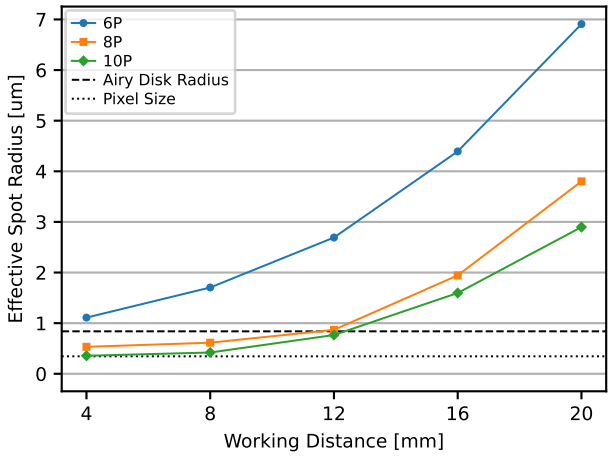
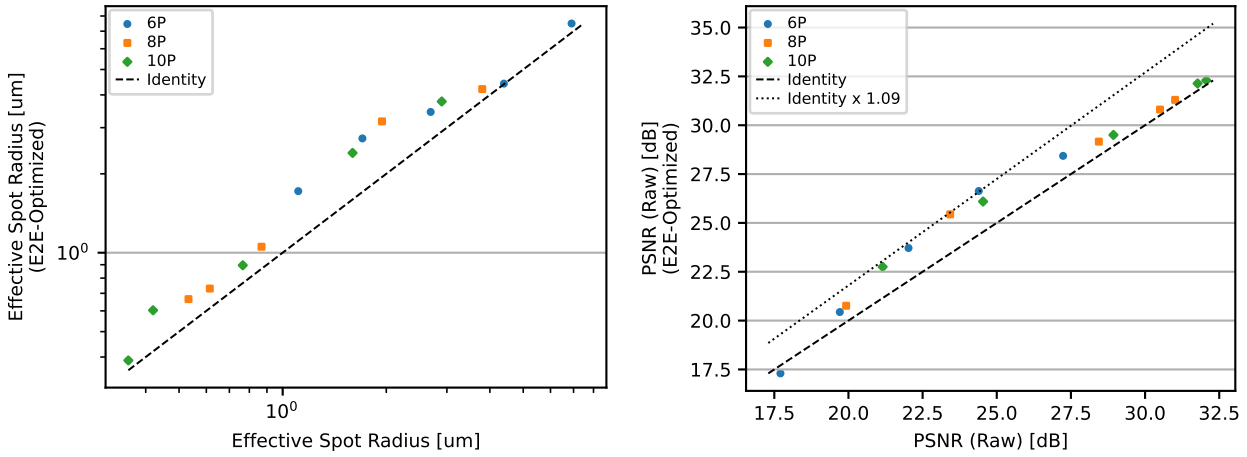


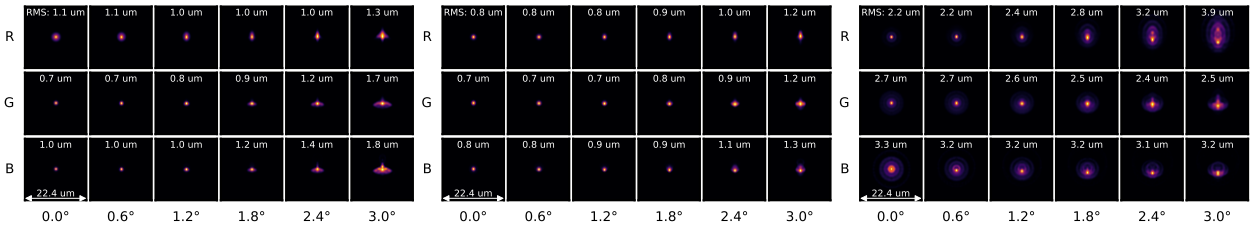
Figure S32: Comparison of spot size for different WDs in the spot-optimized scenario. The displayed pixel size (0.345 µm) corresponds to the object space equivalent of the actual pixel size (considering 20× magnification).



(a) ESR

(b) Raw PSNR

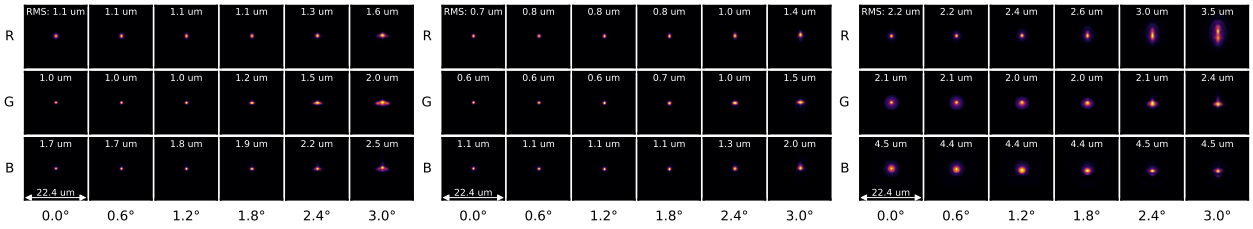
Figure S33: Comparison of ESR (Figure S33a) and raw PSNR (Figure S33b) for the MSE- and spot-optimized scenarios.



(a) Spot-Opt., 6P, WD = 4 mm

(b) Spot-Opt., 8P, WD = 12 mm

(c) Spot-Opt., 10P, WD = 20 mm



(d) MSE-Opt., 6P, WD = 4 mm

(e) MSE-Opt., 8P, WD = 12 mm

(f) MSE-Opt., 10P, WD = 20 mm

Figure S34: Comparisons of PSF for the MSE- and spot-optimized scenarios.

S6 End-to-End Design of C-Mount Camera

In this section, we present the task-driven design of a C-mount camera lens, comparing three different optimization strategies: spot size optimization, image-driven optimization for raw image quality, and end-to-end (E2E) optimization for post-restoration image quality.

In Section S6.1, we provide a detailed formulation of the problem. In Section S6.2, we present the results of these optimization strategies.

S6.1 Problem Formulation

We consider the design of a C-mount camera lens, commonly used in machine vision and scientific imaging applications. The C-mount standard specifies a flange focal distance of 17.526 mm. Our goal is to design a lens with a focal length of 28 mm and an f-number of 2.8, based on the GS3-U3-41C6C-C 1″ FLIR Grasshopper®3 High Performance USB 3.0 Color Camera equipped with a CMOIS CMV4000 sensor. The sensor has a diagonal of 15.93 mm, a pixel pitch of 5.5 µm, and a resolution of 2048×2048. The chosen focal length of 28 mm corresponds to a field of view (FOV) of 31.8°. We set the f-number to 2.0.

Lens configuration We consider a 4-element lens configuration with spherical surfaces only, with the aperture stop positioned in front of the lens. As a result, we do not allow for vignetting, unlike in the telephoto lens study.

Spectrum The camera's quantum efficiency data is used to weight the wavelengths in the optimization objective, as illustrated in Figure S35.

Optimization objective In addition to TRA/GTRA residuals, as discussed in Table S11 and Section S5.1.3, we incorporate soft constraints to ensure the design meets several key requirements: angles of incidence and refraction are kept below 60°, the edge-to-thickness ratio is between $1/3$ and 3, airspaces are greater than 0.25 mm, element thickness is between 1 mm and 5 mm, image clearance is greater than 20 mm (to satisfy the flange focal distance), distortion is less than 2%, and realistic glass materials are optimized for. Glass material selection follows the same catalog and strategy as outlined in Section S5.1.3.

Solves To achieve the desired focal length, we apply a curvature solve on the last refractive surface.

S6.2 Results and Analysis

Methodology We borrow the methodology from the telephoto lens and microscope objective lens case studies. We first optimize the lens for effective spot radius (ESR), obtaining the baseline design displayed in Figure S36a. We then optimize the lens in the *image-driven* setting, that is, for raw image quality (Figure S36c). Finally, we optimize the lens in the *E2E image restoration* setting for restored image quality (Figure S36e) alongside the same image restoration model (IRM) architecture

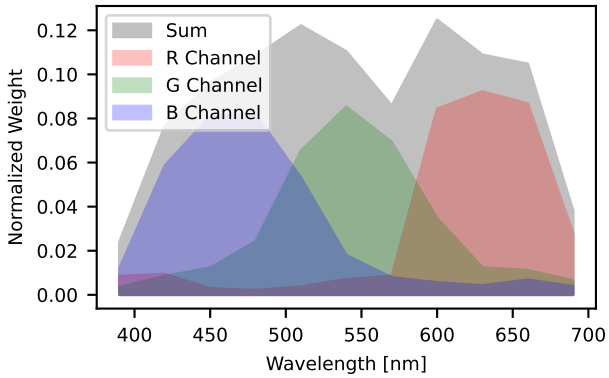


Figure S35: We use the quantum efficiency curves of the FLIR camera to define the spectrum for the C-mount lens design. The spectrum is represented by a regular grid of 11 wavelengths between 390–690 nm. The monochromatic spectrum, obtained by summing the RGB channel spectra, is used in optimizing the lens for effective spot radius (ESR).

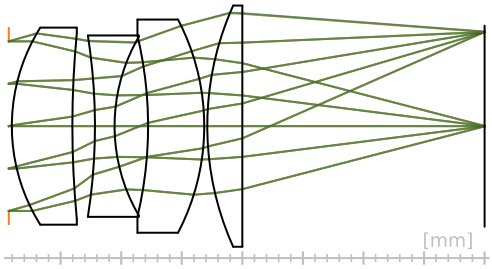
as for telephoto lenses (Section S3.2). For E2E optimization and for validation, we use the DIV2K dataset [1] in the same setting as for the telephoto lens study. The results are summarized in Table S15.

Table S15: Comparison of raw and restored image quality for different lens optimizations.

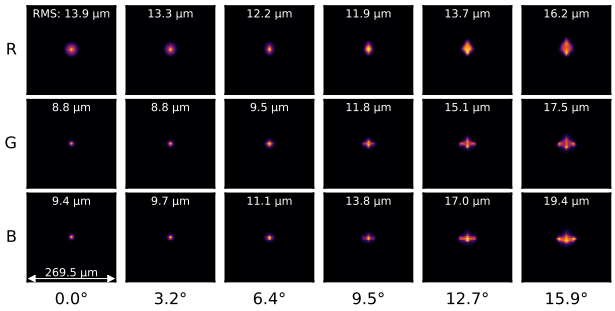
Optimization target	**ESR** [µm]	**Pre-restoration** **PSNR** [dB]	**SSIM**	**Post-restoration** **PSNR** [dB]	**SSIM**
Effective spot radius	12.9	26.2	0.745	34.3	0.934
Raw image quality	15.5 (+2.6)	26.9 (+0.7)	0.784 (+0.039)	36.5 (+2.2)	0.957 (+0.023)
Restored image quality	18.5 (+5.6)	26.3 (+0.1)	0.763 (+0.018)	37.3 (+3.0)	0.963 (+0.029)

PSF comparison We compare the point spread functions (PSFs) of the three settings in Figure S36. Consistent with the other case studies, the PSFs of MSE-optimized lenses exhibit narrower and rounder central peaks compared to the spot-optimized lens, though at the cost of extended tails. This effect is particularly noticeable in the restored image quality setting, where the ESR is significantly higher, with a value of 18.5 µm, compared to 12.9 µm for the spot-optimized lens. In contrast, the raw image quality setting results in an ESR of 15.5 µm, placing it roughly midway between the spot-optimized and restored image quality settings.

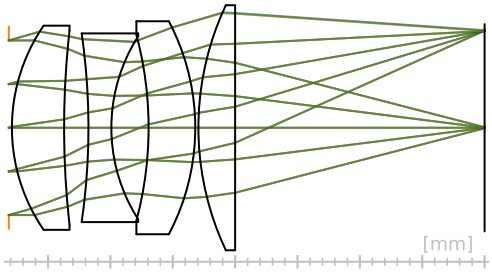
Raw image quality comparison As anticipated, the pre-restoration image quality on the validation set is highest when the lens is optimized for pre-restoration image quality during training. The pre-restoration peak signal-to-noise ratio (PSNR) for the raw image quality-optimized lens is 26.9 dB, and the structural similarity index (SSIM) is 0.784, both surpassing the spot-optimized lens, which has a pre-restoration PSNR of 26.2 dB and SSIM of 0.745. When the lens is optimized for restored image quality, the pre-restoration image quality still outperforms the spot-optimized



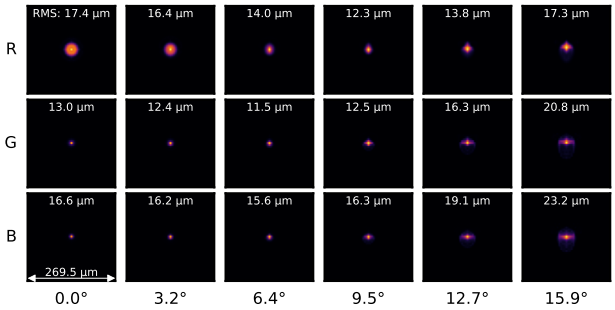
(a) Layout, Effective spot radius



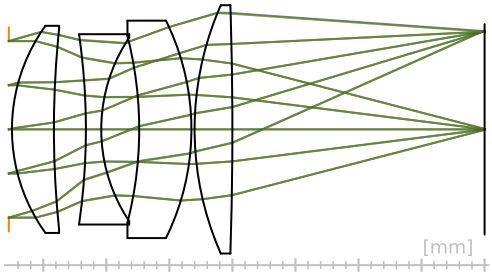
(b) PSFs, Effective spot radius



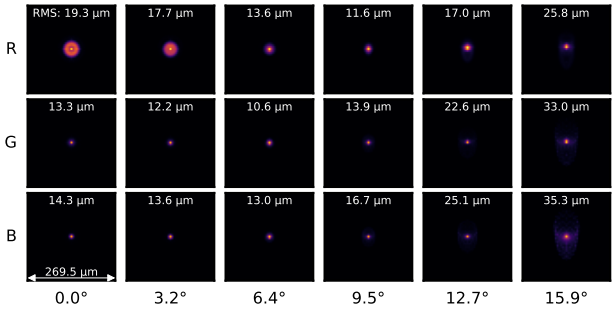
(c) Layout, Raw Image Quality



(d) PSFs, Raw Image Quality



(e) Layout, Restored Image Quality



(f) PSFs, Restored Image Quality

Figure S36: 2D layouts and PSFs for the C-mount camera lens are displayed for three optimization approaches: the spot-optimized lens (Figures S36a and S36b), the MSE-optimized lens focused on raw image quality (Figures S36c and S36d), and the E2E-optimized lens enhanced for restored image quality (Figures S36e and S36f).

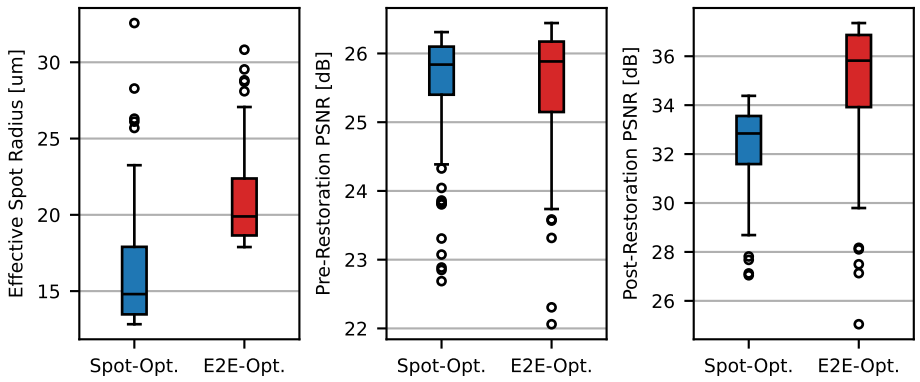


Figure S37: Tolerancing analysis for the spot-optimized and the restored image quality-optimized lenses.

lens, though the difference is less pronounced. The restored image quality-optimized lens achieves a pre-restoration PSNR of 26.3 dB and SSIM of 0.763, values that are closer to the pre-restoration image quality-optimized lens, especially in terms of SSIM.

Restored image quality comparison As expected, when the lens is optimized for restored image quality during training, the performance on the validation set is highest for restored image quality. The restored image quality-optimized lens achieves a restored PSNR of 37.3 dB and SSIM of 0.963. In comparison, the raw image quality-optimized lens achieves a restored PSNR of 36.5 dB and SSIM of 0.957. Both restored values for the restored image quality-optimized lens exceed those of the spot-optimized lens, which has a restored PSNR of 34.3 dB and SSIM of 0.934.

Tolerancing analysis We perform a tolerancing analysis on the restored image quality-optimized lens, as shown in Figure S37. Our Monte-Carlo tolerancing analysis for C-mount lenses uses the following tolerance values: ±0.02 mm for airspaces, ±0.01 mm for element thicknesses, ±0.1% for curvatures, ±0.0005 for refractive indices, and ±0.5% for Abbe numbers. For each perturbed lens (n=100), we apply random perturbations sampled uniformly from these tolerance distributions, then evaluate the average pre- and post-restoration PSNR across all perturbed lenses. The results validate that, for the majority of perturbed lenses, the margins for post-restoration image quality remain well preserved.

S7 Miscellaneous Analyses

S7.1 PSFs of Wide-Angle Smartphone Lens

We extend the point spread function (PSF) validation of Sec. 5.4 (main manuscript) to the 6-element wide-angle smartphone lens optimized in Sec. 5.3 (main manuscript), which has an 82.8° field of view (FOV). Figure S38 compares our coherent Airy disk convolution PSFs against diffraction PSFs from CODE V across 11 field points (0–41.4°) and three wavelengths (450, 550, and 650 nm). Our method achieves an average NRMSE of 6.9 %. Because this lens operates well above the diffraction limit for higher field values, geometric aberrations dominate the PSF shape and all three estimation methods yield similar NRMSE values (geometric: 7.2 %, naive convolution: 7.8 %, coherent: 6.9 %), in contrast to the near-diffraction-limited telephoto lens where the coherent method provides a clear advantage (4.2 % vs. 15.5 % for the geometric method).

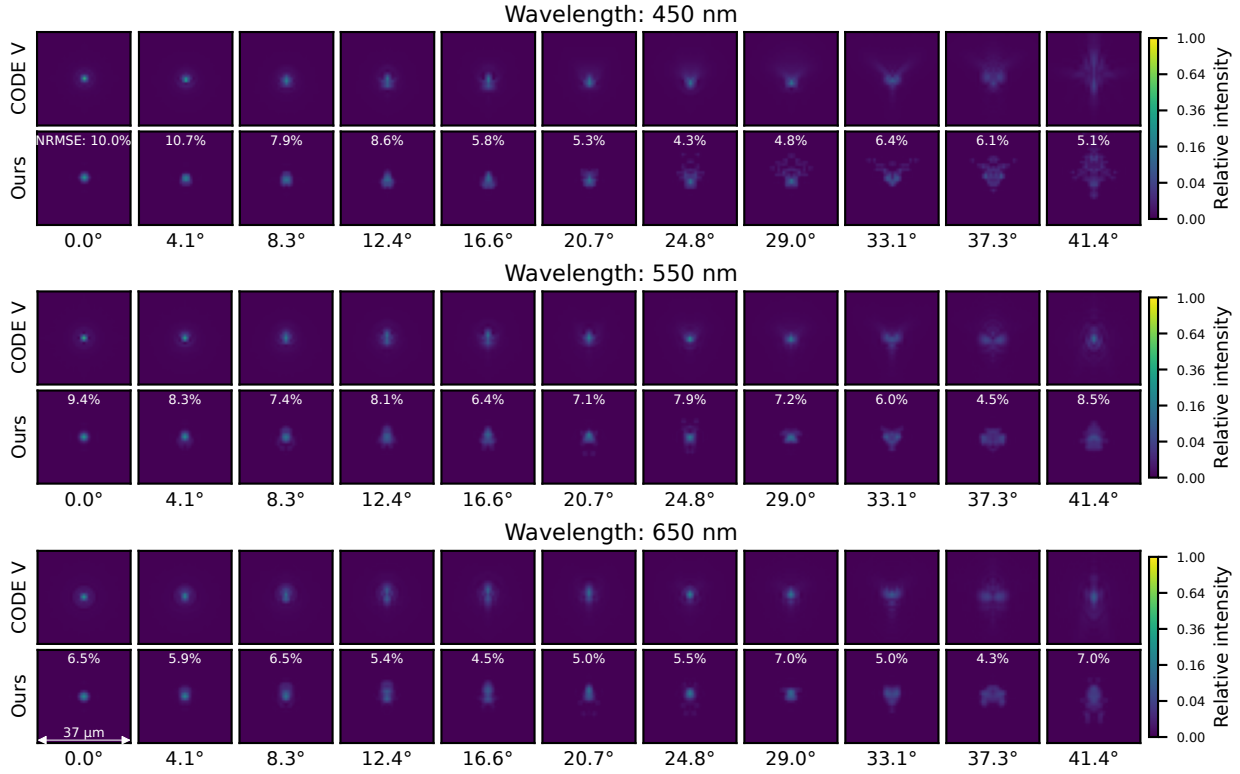


Figure S38: **PSF comparison for the wide-angle smartphone lens (**82.8° **FOV).** For each wavelength (450, 550, and 650 nm), the top row shows the CODE V diffraction PSF and the bottom row shows our coherent Airy disk convolution estimate. Columns span 11 field points from 0° (on-axis) to 41.4° (full field). NRMSE per field is annotated on each bottom row. The square root of intensity is displayed to enhance tail visibility.

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.

[2] Léon Bottou. Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'2010: 19th International Conference on Computational StatisticsParis France, August 22-27, 2010 Keynote, Invited and Contributed Papers*, pages 177–186. Springer, 2010.

[3] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *European Conference on Computer Vision*, pages 17–33. Springer, 2022.

[4] Geoffroi Côté, Jean-François Lalonde, and Simon Thibault. Deep learning-enabled framework for automatic lens design starting point generation. *Opt. Express*, 29(3):3841–3854, Feb 2021. doi: 10.1364/OE.401590.

[5] Geoffroi Côté, Yueqian Zhang, Christoph Menke, Jean-François Lalonde, and Simon Thibault. Inferring the solution space of microscope objective lenses using deep learning. *Optics Express*, 30(5):6531–6545, February 2022. ISSN 1094-4087. doi: 10.1364/OE.451327.

[6] 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.

[7] Alexandre Cléroux Cuillerier, Jeck Borne, and Simon Thibault. Fast metasurface hybrid lens design using a semi-analytical model. *JOSA B*, 40(1):72–78, 2023.

[8] Jiangxin Dong, Stefan Roth, and Bernt Schiele. Deep wiener deconvolution: Wiener meets deep learning for image deblurring. *Advances in Neural Information Processing Systems*, 33: 1048–1059, 2020.

[9] Thomas Eboli, Jian Sun, and Jean Ponce. End-to-end interpretable learning of non-blind image deblurring. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVII 16*, pages 314–331. Springer, 2020.

[10] Greg W Forbes. Shape specification for axially symmetric optical surfaces. *Optics express*, 15 (8):5218–5226, 2007.

[11] Michael Hirsch, Suvrit Sra, Bernhard Schölkopf, and Stefan Harmeling. Efficient filter flow for space-variant multiframe blind deconvolution. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 607–614. IEEE, 2010.

[12] Qi Jiang, Shaohua Gao, Yao Gao, Kailun Yang, Zhonghua Yi, Hao Shi, Lei Sun, and Kaiwei Wang. Minimalist and high-quality panoramic imaging with psf-aware transformers. *arXiv preprint arXiv:2306.12992*, 2023.

[13] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. In *Proccedings of the 3rd International Conference on Learning Representations*, 2015.

[14] Stéphane Larouche and David R Smith. Reconciliation of generalized refraction with diffraction theory. *Optics letters*, 37(12):2391–2393, 2012.

[15] Bruno Lecouat, Thomas Eboli, Jean Ponce, and Julien Mairal. High dynamic range and super-resolution from raw image bursts. *ACM Transactions on Graphics (TOG)*, 41(4):1–21, 2022.

[16] Der-Tsai Lee and Bruce J Schachter. Two algorithms for constructing a delaunay triangulation. *International Journal of Computer & Information Sciences*, 9(3):219–242, 1980.

[17] HW Lee. The hartmann formula for the dispersion of glass. *Transactions of the Optical Society*, 28(3):161, 1926.

[18] Kenneth Levenberg. A method for the solution of certain non-linear problems in least squares. *Quarterly of applied mathematics*, 2(2):164–168, 1944.

[19] Donald W Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the society for Industrial and Applied Mathematics*, 11(2):431–441, 1963.

[20] Yurii Nesterov. A method for unconstrained convex minimization problem with the rate of convergence o (1/k2). In *Dokl. Akad. Nauk. SSSR*, volume 269, page 543, 1983.

[21] Jorge Nocedal and Stephen J Wright. Numerical optimization 2nd edition, 2006.

[22] Matthew P Rimmer. Relative Illumination Calculations. In Robert E. Fischer and Philip J. Rogers, editors, *Optical System Design, Analysis, Production for Advanced Technology Systems*, volume 0655, pages 99 – 104. International Society for Optics and Photonics, SPIE, 1986. doi: 10.1117/12.938414.

[23] 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.

[24] Emmanuel Rousseau and Didier Felbacq. Concept of a generalized law of refraction: a phenomenological model. *ACS photonics*, 7(7):1649–1654, 2020.

[25] Sebastian Ruder. An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*, 2016.

[26] Markus Schake. Examining and explaining the "generalized laws of reflection and refraction" at metasurface gratings. *JOSA A*, 39(8):1352–1359, 2022.

[27] SCHOTT. Refractive index and dispersion. SCHOTT Technical Information TIE-29, 2016.

[28] Rainer Schuhmann. Description of aspheric surfaces. *Advanced Optical Technologies*, 8(3-4): 267–278, 2019.

[29] Siyuan Shen, Zhaohui Ruan, Yuan Yuan, and Heping Tan. Conditions for establishing the "generalized snell's law of refraction" in all-dielectric metasurfaces: theoretical bases for design of high-efficiency beam deflection metasurfaces. *Nanophotonics*, 11(1):21–32, 2021.

[30] Warren J. Smith. *Modern Lens Design*. McGraw Hill Professional, November 2004. ISBN 978-0-07-177726-1.

[31] Qilin Sun, Congli Wang, Qiang Fu, Xiong Dun, and Wolfgang Heidrich. End-to-end complex lens design with differentiate ray tracing. *ACM Trans. Graph.*, 40(4):1–13, jul 2021. ISSN 0730-0301. doi: 10.1145/3450626.3459674.

[32] Synopsys. Code V 2024.03 Documentation Library. Technical report, Synopsys, 2024.

[33] Mark K Transtrum and James P Sethna. Improvements to the levenberg-marquardt algorithm for nonlinear least-squares minimization. *arXiv preprint arXiv:1201.5885*, 2012.

[34] Ethan Tseng, Shane Colburn, James Whitehead, Luocheng Huang, Seung-Hwan Baek, Arka Majumdar, and Felix Heide. Neural nano-optics for high-quality thin lens imaging. *Nature communications*, 12(1):1–7, 2021.

[35] Michael Unser, Philippe Thevenaz, and Leonid Yaroslavsky. Convolution-based interpolation for fast, high-quality rotation of images. *IEEE Transactions on image processing*, 4(10):1371–1381, 1995.

[36] Daniel Werdehausen, Sven Burger, Isabelle Staude, Thomas Pertsch, and Manuel Decker. Dispersion-engineered nanocomposites enable achromatic diffractive optical elements. *Optica*, 6(8):1031–1038, 2019.

[37] Nanfang Yu, Patrice Genevet, Mikhail A Kats, Francesco Aieta, Jean-Philippe Tetienne, Federico Capasso, and Zeno Gaburro. Light propagation with phase discontinuities: generalized laws of reflection and refraction. *science*, 334(6054):333–337, 2011.

[38] Bo Zhang, Josiane Zerubia, and Jean-Christophe Olivo-Marin. Gaussian approximations of fluorescence microscope point-spread function models. *Applied optics*, 46(10):1819–1829, 2007.