Generalized Aberrations for Processing-Aware Optical Design

GEOFFROI CÔTÉ, Princeton University, USA
ETHAN TSENG, Princeton University, USA
FELIX HEIDE, Princeton University, USA

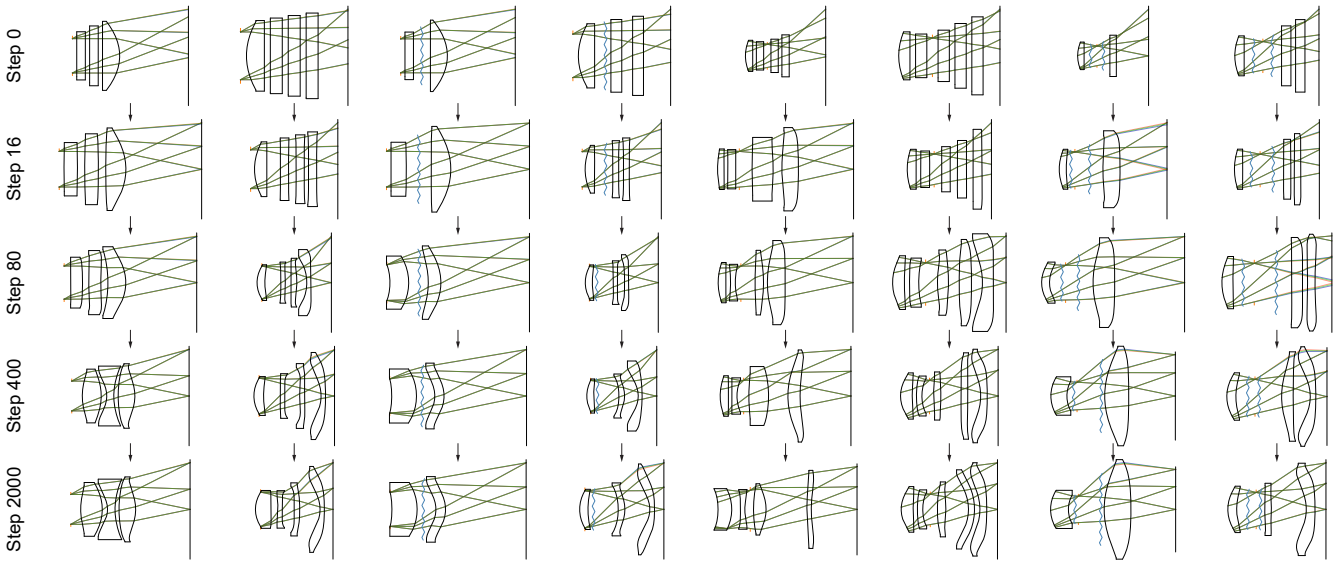

Fig. 1. **Foundational optimization robustness.** Our optimization framework relies on a robust Levenberg-Marquardt (LM) engine designed for complex optical landscapes. We demonstrate its stability by generating designs from rudimentary starts (top row) via conventional spot-radius optimization, accounting for glass materials for hybrid lenses featuring both aspheric and diffractive surfaces (shown as wavy lines). This foundational robustness favors successful outcomes even when optimizing for intricate processing-aware objectives. An extended, animated version is available in the Supplementary Video.

Modern cameras increasingly rely on AI-driven processing, establishing end-to-end (E2E) optimization as a key paradigm in applications ranging from smartphone photography to biomedical imaging. Processing-aware optical design—optimizing the optics for downstream processing—is central to E2E pipelines, yet typically relies on ineffective first-order optimization. Robust pseudo-second-order methods from conventional design, such as Levenberg-Marquardt, generally fail in this context: because such design often relies on a single scalar-valued loss function, it provides insufficient constraints for the high-dimensional optical parameter space, leading to rank-deficient Jacobians that break conventional least-squares solvers.

To address this gap, we introduce the generalized transverse ray aberrations (GTRA): a generalization of the well-established TRA formulation underlying industry-standard spot-radius optimization. By lifting scalar task-driven losses into high-dimensional ray-level residuals, the GTRA objective unlocks Levenberg-Marquardt solvers to bring together the robustness of traditional design with the flexibility of E2E pipelines.

We validate our method on over 100 design instances for E2E image restoration, including smartphone telephoto lenses, microscope objectives, and C-mount cameras. Our optimization decisively outperforms first-order methods, yielding designs that consistently surpass spot-radius-optimized counterparts in image quality or form factor. These findings demonstrate that the GTRA framework extends the robustness of conventional design to the processing-aware setting.

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

Additional Key Words and Phrases: End-to-End Optical Design, Image Restoration, Computational Imaging, Computer-Aided Design

ACM Reference Format:
Geoffroi Côté, Ethan Tseng, and Felix Heide. 2026. Generalized Aberrations for Processing-Aware Optical Design. *ACM Trans. Graph.* X, X, Article X (May 2026), 17 pages. https://doi.org/10.1145/3817055

Authors' addresses: Geoffroi Côté, Princeton University, USA, gcote@princeton.edu; Ethan Tseng, Princeton University, USA, eftseng@princeton.edu; Felix Heide, Princeton University, USA, fheide@princeton.edu.

ACM 0730-0301/2026/5-ARTX
https://doi.org/10.1145/3817055

1 INTRODUCTION

Despite their diverse uses, most modern imaging systems share a common process: capturing light through an optical system and processing the resulting image. The performance of imaging optics has traditionally been assessed using criteria such as effective spot radius, enabling designers to optimize optical systems in isolation without simulating the actual imaging process [Bentley and Olson 2012; Kidger 2004; Smith 2004]. However, the growing reliance on

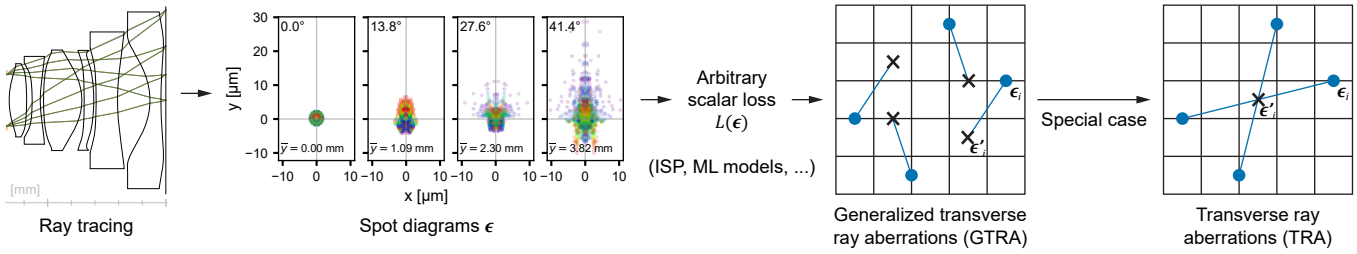

Fig. 2. **Generalized transverse ray aberrations (GTRA) enable effective processing-aware optimization of complex imaging optics.** The scalar nature of typical loss functions for end-to-end (E2E) co-design tasks excludes effective optics optimization with the LM algorithm. We introduce GTRA as a method to lift any differentiable scalar loss into the intermediate, high-dimensional spot-diagram space, enabling its use with least-squares solvers. The GTRA objective generalizes transverse ray aberrations (TRA), commonly employed for conventional lens optimization, to more general contexts.

data-driven models for image restoration and other computer vision tasks, notably in smartphone photography and medical applications, has elevated the role of processing in optical imaging [Barbastathis et al. 2019; Wetzstein et al. 2020]. Traditional performance criteria fail to capture this expanding interplay between physical image formation and digital processing, prompting a reevaluation of the siloed design paradigm.

Task-driven optical design addresses this challenge by modeling the optics and processing together, and optimizing the optics for a task-specific loss function, typically scalar-valued. This approach is particularly relevant in *end-to-end* (E2E) *design*, where the optical components and downstream model are jointly optimized for a common objective. By sharing the burden of correcting aberrations with the data-driven model, the lens can focus on capturing the scene's most critical information. E2E design has been successfully applied to various downstream tasks; in this work, we focus on *processing-aware optical design*—the optimization of imaging optics for a downstream image-processing pipeline—exemplified by *E2E image restoration*, where the optics are optimized alongside an image restoration model (IRM). This method aims to produce optics that maintain the high image quality of traditional optics while being more cost-effective, lightweight, or compact.

Despite its theoretical benefits, E2E design introduces a unique set of challenges. Beyond the complexities inherent to conventional optical design [Smith 2004], these challenges include noisy and intricate downstream loss functions, the joint optimization of processing, and the impracticality of human intervention [Côté et al. 2023a,b]. Most existing task-driven design approaches [Côté et al. 2023a; Sun et al. 2021; Tseng et al. 2021a,b; Yang et al. 2024a, 2023] rely on stochastic gradient descent (SGD) due to its synergy with deep learning [LeCun et al. 2015] and compatibility with automatic differentiation (AD) platforms [Abadi et al. 2016; Paszke et al. 2019]. However, SGD often struggles with the highly non-linear optimization landscapes of optical designs [Gross 2007], which are riddled with local minima and saddle points [Turnhout and Bociort 2009]. As a result, existing E2E works have either focused on low-complexity problem spaces—with variables few in number [Côté et al. 2023a; Sun et al. 2021; Tseng et al. 2021a; Yang et al. 2023] or artificially bounded [Tseng et al. 2021b]—or yielded markedly suboptimal solutions [Yang et al. 2024a] (see Sec. 5.3).

In contrast, industry-standard *conventional* optical design software [Ansys, Inc. 2025; Keysight, Inc. 2025] has long relied on the Levenberg-Marquardt (LM) algorithm [Levenberg 1944; Marquardt 1963], framing the problem as a least-squares one and leveraging the Jacobian for effective optimization. However, this powerful machinery cannot be directly applied to task-driven design. Unlike conventional design which optimizes thousands of ray aberrations, task-driven design typically minimizes a single scalar-valued loss function. This creates a fundamental dimensionality mismatch: a scalar loss lacks the degrees of freedom to define a full-rank Jacobian for the high-dimensional optical parameter space, rendering standard LM solvers inapplicable [Nocedal and Wright 2006].

In this work, we introduce the first, to our knowledge, method that bridges this dimensionality gap, advancing LM optimization from conventional to processing-aware design. Our approach introduces the *generalized transverse ray aberrations* (GTRA), a least-squares objective that "lifts" the scalar losses typical of data-driven pipelines into ray-level residuals compatible with LM, as illustrated in Fig. 2. GTRA extend the well-established transverse ray aberrations (TRA) used for spot-radius optimization [Keysight, Inc. 2025]—and reduce to TRA for conventional design, as in Fig. 1—bringing the robustness of traditional least-squares solvers to the processing-aware setting.

We validate the effectiveness of our approach on E2E image restoration (Fig. 3) across diverse optical configurations, including smartphone telephoto, microscope objective, and C-mount lenses. *We demonstrate that our GTRA-based optimizer outperforms first-order methods in both convergence speed and solution quality. In over 100 design instances, our processing-aware optimization consistently yields lenses that surpass conventionally designed counterparts in image quality despite their larger spot radii.* These improvements are partly explained by the optimizer's ability to exploit PSFs with narrow central peaks and elongated tails—features penalized by conventional spot-radius metrics but beneficial for restoration. Equivalently, our E2E designs achieve comparable image quality with fewer elements, reduced total track length (TTL) in smartphone telephoto lenses, or increased working distance (WD) in microscope objective lenses. More broadly, these results reveal the significant potential of processing-aware optimization: by moving beyond conventional spot-radius metrics, the optimizer discovers design trade-offs that better serve the end application—a capability that spot-radius optimization, by construction, cannot provide.

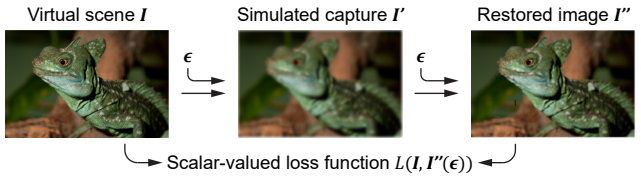

Fig. 3. **E2E image restoration.** We address E2E image restoration by jointly optimizing the lens design and image restoration model (IRM), minimizing the difference between the original I and restored I'' images. Spot diagrams ϵ are a key quantity in computing the downstream loss function $L(I, I''(\epsilon))$, as they serve as an intermediate representation linking the lens variables to the image simulation and restoration processes.

Beyond the GTRA formulation and its validation, we contribute several practical capabilities motivated by real-world lens design: diffraction-compensated geometric PSFs, mesh-based constraints for material selection and optimization, and optimizable vignetting. These capabilities address practical design constraints often overlooked in computational imaging literature, making our E2E optics optimization applicable to real-world engineering workflows.

Our code and designs are available at https://light.princeton.edu/generalized-aberrations.

Scope and Limitations. The derivation of GTRA assumes that lens variables influence the downstream loss function through an intermediate representation: the spot diagrams. In settings where spot diagrams are not explicitly computed—for instance, wave-optics simulations, detector-to-scene ray tracing [Jakob et al. 2022], or full diffraction treatments—this assumption may require additional handling. Nonetheless, dependencies beyond the spot diagrams can still be incorporated as additional residuals in the least-squares formulation.

While the GTRA formulation is mathematically agnostic to the choice of scalar loss, our validation focuses on *processing-aware optical design*—specifically E2E image restoration. Task-agnostic generalizability to higher-level perception tasks (e.g., object detection, semantic segmentation) remains unvalidated and is left as future work. Our aim, accordingly, is to advance the optimization of imaging optics within current design pipelines rather than to address physical manufacturing or prototyping; the validation we present focuses on optimization robustness, demonstrating consistent convergence across over 100 design instances. Although orthogonal to the optimization contribution, we additionally verify simulation accuracy against industry-standard software (CODE V [Keysight, Inc. 2025]) and confirm through Monte-Carlo tolerancing that E2E-optimized designs are not more sensitive to manufacturing variations than their conventionally optimized counterparts. Investigating the remaining sim-to-real gap through physical fabrication is a promising direction for future work.

2 RELATED WORK

Conventional Optics Optimization. Traditional optical design begins with a set of required specifications that define the optimization problem. The design configuration, determined by the sequence of optical elements—e.g., spherical or aspheric lenses, or diffractive optical elements (DOEs)—defines the possible variables, such as surface curvatures and spacings. In conventional design, the optimization objective is often formulated as a least-squares problem, or merit function, with residuals that include TRA or other performance targets, as well as various soft constraints [Bentley and Olson 2012]. The LM algorithm [Girard 1958; Wynne 1959], a non-linear least-squares method, is the established local optimizer for lens design and serves as the default optimizer in major optical design packages [Ansys, Inc. 2025; Keysight, Inc. 2025]. We find our work to provide a bridge between conventional and task-driven optics optimization.

Task-Driven Design. Task-driven optical design extends conventional design by modeling both the image capture and processing stages, and optimizing the optics for a task-specific loss function. This plays a fundamental role in E2E design pipelines where processing components—e.g., the image signal processing (ISP) [Tseng et al. 2021b] or downstream deep neural networks—are optimized alongside the optics. Image restoration is a key task in E2E design, naturally aligning with the goal of conventional lens design: maximizing the overall image quality of the imaging system. Subtasks include high-dynamic-range imaging [Metzler et al. 2020; Sun et al. 2020], large field-of-view imaging [Peng et al. 2019], extended depth-of-field imaging [Sitzmann et al. 2018; Yang et al. 2024a; Zhang et al. 2023a], and super-resolution imaging [Sitzmann et al. 2018]. Many perception tasks have also been explored, such as image classification [Chang et al. 2018], depth estimation [Chang and Wetzstein 2019; Haim et al. 2018], and object detection [Chang and Wetzstein 2019; Côté et al. 2023a; Tseng et al. 2021b]. While the GTRA formulation applies mathematically to any downstream task, we focus on processing-aware optical design—specifically E2E image restoration—a challenging benchmark for measuring the benefits of E2E over conventional design.

SGD Optimization and Limitations. As many task-driven problems originate in deep learning, performance is commonly evaluated using a scalar-valued loss and optimized with SGD or adaptive variants such as Adam [Kingma and Ba 2015], which are effective at training high-dimensional neural networks and thus widely adopted in E2E optics works. However, SGD has well-known limitations when applied to optics optimization [Gross 2007]: it converges slowly, can get stuck in poor local minima, and—by reducing the loss to a scalar—cannot exploit the structure that pseudo-second-order methods like LM use to accelerate convergence. In particular, SGD variants lack scale invariance [Holl et al. 2022], making them sensitive to parameter scaling, exacerbating the non-linearity of high-degree aspherics, and further slowing or destabilizing convergence—a problem some works address by manually adjusting learning rates per polynomial degree [Yang et al. 2024a]. Moreover, E2E pipelines compound these issues because the loss combines multiple complex, dynamically changing components (imaging simulation, ISP, downstream networks) [Côté et al. 2023b]. These observations motivate our approach of converting near-arbitrary scalar objectives into equivalent least-squares forms via the GTRA objective, allowing E2E pipelines to take advantage of established pseudo-second-order optimization techniques from conventional optics design.

We note E2E approaches that rely on existing commercial software for both simulation and optimization [Fontbonne et al. 2022; Stork and Robinson 2008]; however, these approaches tend to overlook the scalar nature of loss functions and do not leverage AD platforms for efficiently computing derivatives of computationally intensive E2E loss functions. Some recent E2E methods explore quasi-global searches to address similar optimization challenges [Gao et al. 2025], and our approach can complement such strategies.

Imaging Simulation. Imaging simulation aims to numerically reproduce the process of image capture, from the virtual scene to the sensor. While some methods [Sun et al. 2021] apply backward (detector-to-scene) ray-tracing simulations, as with Mitsuba 3 [Jakob et al. 2022], the typical method involves evaluating the PSF of the optical system and convolving it over the virtual scene. This is commonly applied in paraxial wave-optics simulations for modeling DOEs [Chang and Wetzstein 2019; Metzler et al. 2020; Sitzmann et al. 2018; Sun et al. 2020] or metasurfaces, where E2E co-design with learned reconstruction has been actively explored [Arya et al. 2024; Chakravarthula et al. 2023; Choi et al. 2024; Fröch et al. 2025; Sun et al. 2025; Tseng et al. 2021a; Zhang et al. 2023b]; see Lee et al. [2024] for a review. In contrast, compound refractive lenses—the design paradigm for most commodity-type cameras—cannot rely on the paraxial approximation in the general case and require exact ray tracing based on Snell's Law. To address this, differentiable ray-tracing frameworks have been introduced in recent years [Côté et al. 2023a; Halé et al. 2021; Li et al. 2021; Sun et al. 2021; Wang et al. 2022] to enable designing compound optics within E2E pipelines, with subsequent works extending this to diffraction PSFs [Yang et al. 2024b; Zhang et al. 2024]. Surrogate models have also been explored [Tseng et al. 2021b], though these restrain the design space to predefined parameter boundaries.

Our imaging simulation model is closely related to that of Côté et al. [2023a], which traces rays from the object to the image plane, obtaining the spot diagrams from which the PSF is computed. However, our proposed method introduces key components that are essential for us to robustly solve generic E2E optics optimization problems. This includes diffraction-compensated PSFs, mesh-based glass constraints, and realistic vignetting modeling and optimization through front apertures.

3 GENERALIZED OPTICAL SYSTEM OPTIMIZATION

We propose an optimization method to bridge the gap between conventional and task-driven design, forming the foundation of our E2E optimization strategy, as detailed next in Sec. 4. We begin by outlining our optimization framework (Sec. 3.1), which builds upon proven methods for conventional optical design and is demonstrated in the Supplementary Video with diverse problems.

Most existing E2E methods rely on first-order optimization for task-driven optical design. However, as our analysis and experiments show, LM optimization with the traditional TRA objective largely outperforms SGD methods in conventional design. Since conventional design is mathematically a restricted (and simpler) subset of task-driven design, this observation motivates the use of pseudo-second-order methods more broadly. As part of this approach, we highlight a fundamental rank deficiency issue that limits the applicability of the LM algorithm when extended to task-driven design (Sec. 3.2). Without resolving this issue, *naively applying LM to task-driven design leads to poor performance* (for example, following the LM-based formulation of Wang et al. [2022]). To illustrate these concepts, we present a toy design problem in Fig. 4 and Table 1, discussed in greater detail in Secs. 3.2 and 3.3.

Table 1. **Effectiveness of the proposed optimization in task-driven optical design.** Various optimization strategies are evaluated on the toy problem shown in Fig. 4. All experiments within each set use the same optimization objective $L(\boldsymbol{\theta})$: effective spot radius (ESR) for conventional design, and MSE between the virtual scene $\boldsymbol{I}$ and its simulated capture $\boldsymbol{I'}$ for task-driven design. With the proposed GTRA, the task-driven lens achieves the lowest MSE by a large margin. This mirrors the effectiveness of TRA in ESR optimization, but applied to the task-driven context. For existing SGD and Adam methods, only the best attempt is reported. See Fig. 4 for convergence plots and illustrated final designs.

	ESR [µm] $\propto \|\|\boldsymbol{\epsilon} - \overline{\boldsymbol{\epsilon}}\|\|$	MSE $\propto \|\|\boldsymbol{I'} - \boldsymbol{I}\|\|^2$
Conventional design, $L \propto \|\|\boldsymbol{\epsilon}(\boldsymbol{\theta}) - \overline{\boldsymbol{\epsilon}}\|\|^2$		
SGD	69.4	–
Adam	32.8	–
LM + $\boldsymbol{\ell}_{\text{ESR}}$ (Eq. (4))	33.0	–
†LM + $\boldsymbol{\ell}_{\text{TRA}}$ (Eq. (3))	**16.6**	0.0100
Image-driven design, $L \propto \|\|\boldsymbol{I'}(\boldsymbol{\theta}) - \boldsymbol{I}\|\|^2$		
SGD [Sitzmann et al. 2018]	–	0.0189
Adam [Côté et al. 2023a; Yang et al. 2024a]	–	0.0080
LM + $\boldsymbol{\ell}_{\text{MSE}}$ (Eq. (7)) [Wang et al. 2022]	–	0.0235
Proposed – LM + $\boldsymbol{\ell}_{\text{GTRA}}$ (Eq. (6))	35.8	**0.0033**

†Industry standard for siloed optical design [Keysight, Inc. 2025]

To tackle this challenge, we introduce GTRA in Sec. 3.3, which enable LM optimization to be effectively extended to processing-aware design problems. We further provide overviews of our ray-tracing and optimization methodologies in Sec. 3.4 and Sec. 3.5, respectively.

3.1 Conventional Optics Optimization

Conventionally, the primary objective in optimizing imaging optics is to maximize optical performance while meeting various design constraints, such as geometry and material compatibility. In the absence of hard constraints, the goal is to minimize the objective function $L(\boldsymbol{\theta})$, that is

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} L(\boldsymbol{\theta}), \tag{1}$$

where the parameterization $\boldsymbol{\theta} \in \mathbb{R}^N$ typically consists of several dozens of heterogeneous variables in complex lenses, including (aspheric) surface profiles, material properties, and spacings between adjacent surfaces. The optimization objective is commonly formulated as the sum-of-squares objective

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

where $\boldsymbol{\ell} \in \mathbb{R}^M$ represents the *residuals*, encompassing multiple optical performance targets and soft constraints. As an essential part of the optimization objective, the ray-tracing process is used to

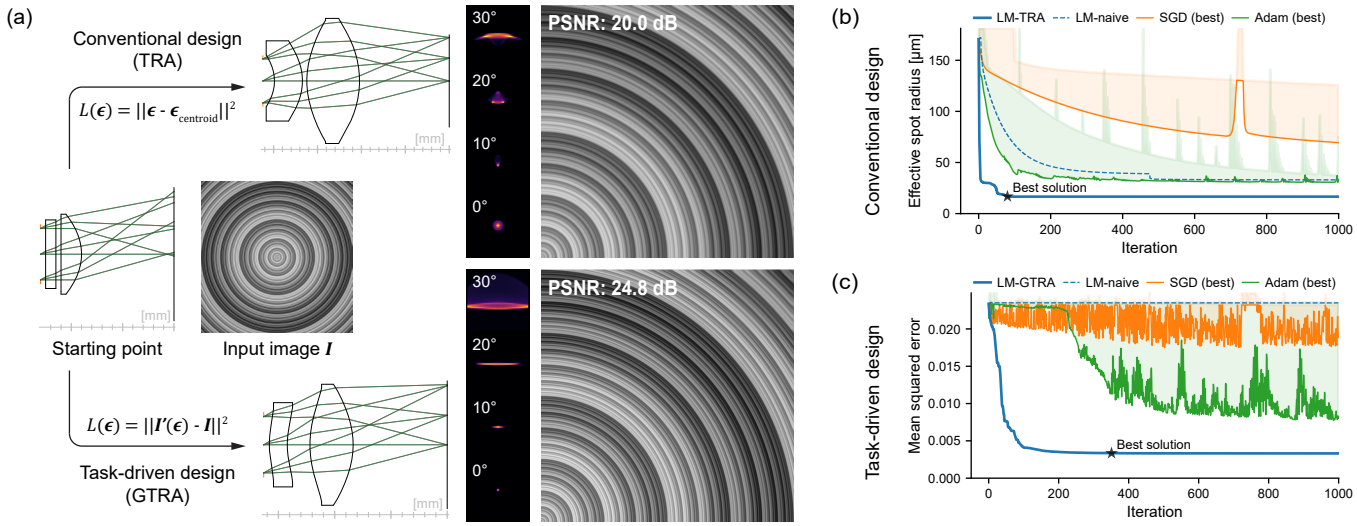

Fig. 4. **Proposed task-driven optimization applied to a simple two-element lens system.** (a) Two optimization scenarios are compared: (1) *conventional*, minimizing spot radius using the established TRA objective, and (2) *task-driven*, minimizing MSE between original and simulated images using the proposed GTRA objective. (b) In scenario 1, our LM algorithm with TRA converges rapidly to a solution that significantly outperforms SGD and Adam, commonly used in E2E works [Côté et al. 2023a; Yang et al. 2024a]. Replacing TRA with a first-order-equivalent 1-dimensional residual slows LM convergence and yields a worse solution. (c) In scenario 2, we observe an even larger margin of improvement for our LM algorithm with GTRA over all alternatives.

compute spot diagrams $\boldsymbol{\epsilon}(\boldsymbol{\theta}) \in \mathbb{R}^{2N_\mathrm{f}N_\mathrm{w}N_\mathrm{p}}$ representing the (x, y) coordinates of the rays at the image plane for N_f different field values, N_w wavelengths, and N_p pupil coordinates. Spot diagrams are a crucial quantity: in conventional design, they are used to compute the effective spot radius, and in task-driven design, they can serve as an intermediate representation of the lens for imaging simulation. In conventional design, a common objective is to minimize TRA at the image plane, defined as

$$\boldsymbol{\ell}_\mathrm{TRA}(\boldsymbol{\theta}) = \frac{1}{\sqrt{N_\mathrm{f}N_\mathrm{w}N_\mathrm{p}}} \left(\boldsymbol{\epsilon}(\boldsymbol{\theta}) - \overline{\boldsymbol{\epsilon}}(\boldsymbol{\theta})\right) \in \mathbb{R}^{2N_\mathrm{f}N_\mathrm{w}N_\mathrm{p}}, \tag{3}$$

where weights are omitted for simplicity, and $\overline{\boldsymbol{\epsilon}}$ (with $2N_\mathrm{f}$ unique values) represents the spot diagram centroid for each field. For a given field, $||\boldsymbol{\ell}_\mathrm{TRA}||$ can be recognized as the effective spot radius.

Conforming to the widely adopted standard in lens design software [Meiron 1965], we employ the LM algorithm to solve the optimization problem. Unlike gradient-based optimization, LM relies on the Jacobian matrix $\mathbf{J} \in \mathbb{R}^{M\times N}$, with the crucial condition that $M \geq N$ for effective optimization.

3.2 Rank Deficiency in Task-Driven Design

We now turn toward *task-driven optics optimization*, encompassing a broader range of tasks such as E2E image restoration (illustrated in Fig. 3 and presented in Sec. 4). These tasks, following deep learning methodology, are typically defined by a *scalar-valued* loss function $L_\mathrm{TD}(\boldsymbol{\theta})$, whereas LM expects a *least-squares objective* with $M \geq N$.

We note that the MSE loss is a special case that can satisfy $M \geq N$; however, evaluating $\mathbf{J}$ is impractical when L_TD involves costly imaging simulations or AI-driven processing [Wang et al. 2022], as its computation scales with M (backward-mode AD) or N (forward-mode) [Baydin et al. 2018], becoming prohibitive in complex systems such as smartphone lenses where $N > 100$ is not uncommon.

Naturally, setting L_TD as the sole residual ($M = 1$) of a least-squares problem fails to satisfy $M \geq N$ when $N > 1$. We briefly revisit conventional design to show why this mismatch in dimensionality is problematic. We consider the 1-dimensional effective spot radius as the sole residual,

$$\boldsymbol{\ell}_\mathrm{ESR}(\boldsymbol{\theta}) = [||\boldsymbol{\ell}_\mathrm{TRA}(\boldsymbol{\theta})||] \in \mathbb{R}^1, \tag{4}$$

which shares the same sum-of-squares objective as $\boldsymbol{\ell}_\mathrm{TRA}$ (Eq. (3)). Using $\boldsymbol{\ell}_\mathrm{ESR}$ ($M = 1$) instead of $\boldsymbol{\ell}_\mathrm{TRA}$ ($M = 2N_\mathrm{f}N_\mathrm{w}N_\mathrm{p}$) causes rank deficiency in $\mathbf{J}$, degrading the performance of optimizers that rely on a full column rank $\mathbf{J}$, such as LM, by discarding per-ray sensitivities. Table 2 summarizes the differences.

Toy Conventional Design Problem. To illustrate the consequences of a rank-deficient $\mathbf{J}$, we optimize a two-element monochromatic lens with 20 variables. The upper portion of Fig. 4a illustrates the TRA-optimized design. As shown in Fig. 4b and Table 1, using $\boldsymbol{\ell}_\mathrm{TRA}$ instead of its 1-dimensional counterpart $\boldsymbol{\ell}_\mathrm{ESR}$ leads the LM optimizer to converge faster and to a better solution. This underscores the importance of preserving per-ray information in $\mathbf{J}$, a principle that also applies to the task-driven version of this experiment, reported in the following Sec. 3.3. Gradient-based optimizers (SGD and Adam) also fail to reach an optimal solution across multiple learning rates (0.01, 0.001, or 0.0001), despite the simplicity of the problem.

Setup. The optics, with an FOV of 60°, an f-number of 2, and a focal length of 10 mm maintained via a last curvature solve, include 5 spacings, 3 curvatures, and 12 aspheric coefficients (degrees 4, 6, and 8). Refractive indices are fixed to 1.5. In addition to $\boldsymbol{\ell}_\mathrm{TRA}$

Table 2. **Comparison of conventional and task-driven optimization in our approach.** To optimize the effective spot radius, the established approach for effective optimization with the LM algorithm is to set TRA as residuals [Keysight, Inc. 2025]. In task-driven design, however, scalar-valued losses do not naturally lend themselves to a similar formulation. GTRA addresses this by converting the scalar loss into a least-squares objective, enabling per-ray sensitivity to be exposed to the LM optimizer, as reflected in the Jacobian size. For conciseness, we use a simplified notation.

	Conventional design	Task-driven design
Example goal	Minimize spot radius	Minimize MSE
Objective function	$L = {}^1/_2\|\|\boldsymbol{\epsilon} - \overline{\boldsymbol{\epsilon}}\|\|^2$	$L = {}^1/_2\|\|\boldsymbol{I'}(\boldsymbol{\epsilon}) - \boldsymbol{I}\|\|^2$
Residuals		
Naive approach	$\boldsymbol{\ell}_{\text{ESR}} = [\|\|\boldsymbol{\epsilon} - \overline{\boldsymbol{\epsilon}}\|\|]$	$\boldsymbol{\ell}_{\text{MSE}} = [\|\|\boldsymbol{I'}(\boldsymbol{\epsilon}) - \boldsymbol{I}\|\|]$
TRA/GTRA	$\boldsymbol{\ell}_{\text{TRA}} = \boldsymbol{\epsilon} - \overline{\boldsymbol{\epsilon}}$	$\boldsymbol{\ell}_{\text{GTRA}} = \sqrt{w}\,(\boldsymbol{\epsilon} - \boldsymbol{\epsilon'})$
Jacobian size		
Naive approach	$1 \times N$	$1 \times N$
TRA/GTRA	$2N_{\text{f}}N_{\text{w}}N_{\text{p}} \times N$	$2N_{\text{f}}N_{\text{w}}N_{\text{p}} \times N$

(effective approach) or $\boldsymbol{\ell}_{\text{ESR}}$ (naive approach), the residuals include soft constraints to enforce airspaces ≥0.5 mm, thickness between 0.5–6 mm, edge-to-thickness ratio between 1:3 and 3:1, angles of incidence/refraction ≤60°, and surface normals ≤30° (see Sec. 3.5).

3.3 Generalized Transverse Ray Aberrations

We consider task-driven design scenarios in which the optimization objective $L_{\text{TD}}(\boldsymbol{\theta})$ is scalar-valued—e.g., an object detection loss or an image dissimilarity metric. We make the assumption that the scalar loss depends on the optical parameters $\boldsymbol{\theta}$ only through the intermediate spot diagrams $\boldsymbol{\epsilon}(\boldsymbol{\theta})$, while noting that auxiliary residuals could be incorporated to account for any additional dependencies. As a reminder, LM updates require two quantities (see Sec. 3.5), the residuals $\boldsymbol{\ell} \in \mathbb{R}^M$ and the Jacobian matrix $\mathbf{J} \in \mathbb{R}^{M \times N}$, but scalar losses lack the required high-dimensional residual structure.

To solve the aforementioned rank deficiency issue, we introduce the GTRA objective (Fig. 2): a vector-valued, sum-of-squares surrogate in the spot-diagram domain that exactly matches the scalar loss's value and local gradient at the current iterate. In other words, we start from a scalar objective but convert it into per-ray residuals in a way that preserves the local task-driven gradient while restoring the dimensionality required for pseudo-second-order updates. Our formulation draws on the idea of "lifting" residuals into higher-dimensional spaces in least-squares optimization [Zach 2014], adapting it to the unique structure of task-driven optical design.

Let $\boldsymbol{\epsilon}_0$, $L_{\boldsymbol{\epsilon}_0}$, and $\nabla L_{\boldsymbol{\epsilon}_0}$ be the spot diagrams, scalar-valued loss, and gradient at a given iteration. We define the GTRA objective by approximating L_{TD} into a sum-of-squares objective,

$$\begin{aligned} L_{\text{TD}}(\boldsymbol{\theta}) &\approx L_{\boldsymbol{\epsilon}_0} + \nabla L_{\boldsymbol{\epsilon}_0}^T\,(\boldsymbol{\epsilon}(\boldsymbol{\theta}) - \boldsymbol{\epsilon}_0) \\ &\approx \frac{1}{2}\underbrace{\frac{||\nabla L_{\boldsymbol{\epsilon}_0}||^2}{2L_{\boldsymbol{\epsilon}_0}}}_{w}\left\|\boldsymbol{\epsilon}(\boldsymbol{\theta}) - \underbrace{\left(\boldsymbol{\epsilon}_0 - 2\frac{L_{\boldsymbol{\epsilon}_0}\nabla L_{\boldsymbol{\epsilon}_0}}{||\nabla L_{\boldsymbol{\epsilon}_0}||^2}\right)}_{\boldsymbol{\epsilon'}}\right\|^2, \end{aligned} \tag{5}$$

where the first approximation is a locally gradient-preserving linearization of the scalar loss in the intermediate space of the spot diagrams, and the second approximation, also locally gradient-preserving, recognizes that $||\boldsymbol{\epsilon}(\boldsymbol{\theta}) - \boldsymbol{\epsilon}_0||$ is small near convergence (see Supp. S3.1 for derivation).

Eq. (5) now has the form of a sum-of-squares objective $L_{\text{TD}}(\boldsymbol{\theta}) \approx {}^1/_2||\boldsymbol{\ell}_{\text{GTRA}}(\boldsymbol{\theta})||^2$ with residuals

$$\boldsymbol{\ell}_{\text{GTRA}}(\boldsymbol{\theta}) = \sqrt{w}\,\left(\boldsymbol{\epsilon}(\boldsymbol{\theta}) - \boldsymbol{\epsilon'}\right) \in \mathbb{R}^{2N_{\text{f}}N_{\text{w}}N_{\text{p}}}. \tag{6}$$

In practice, we compute $\mathbf{J}$ using forward-mode AD, treating both w and $\boldsymbol{\epsilon'}$ as constants, and apply LM updates using $\boldsymbol{\ell}_{\text{GTRA}}$ and $\mathbf{J}$.

Relation to Conventional TRA. GTRA generalize TRA. Expanding $L(\boldsymbol{\theta}) = {}^1/_2||\boldsymbol{\ell}_{\text{TRA}}(\boldsymbol{\theta})||^2$ through Eq. (5) recovers TRA as the special case with $w = {}^1/_{N_{\text{w}}N_{\text{p}}N_{\text{f}}}$ and $\boldsymbol{\epsilon'} = \overline{\boldsymbol{\epsilon}}$. Unlike TRA, which aim rays at fixed centroids $\overline{\boldsymbol{\epsilon}}$, GTRA update per-ray targets $\boldsymbol{\epsilon'}$ and weight w each iteration to align the constructed residuals with the local gradient of the scalar task loss (Fig. 2). Interestingly, if $\boldsymbol{\epsilon'}$ (GTRA) and $\overline{\boldsymbol{\epsilon}}$ (TRA) are treated as constants, the two Jacobians coincide up to a scalar factor, so GTRA retain the per-ray sensitivity that underlies TRA's empirical success. Importantly, the iterative retargeting and reweighting in GTRA redistribute the scalar loss gradient across ray-level residuals, enabling LM updates to leverage per-ray information for faster, task-aligned convergence.

Derivatives. We compute the Jacobian matrix of $\boldsymbol{\ell}_{\text{GTRA}}(\boldsymbol{\theta})$ using forward-mode AD [Paszke et al. 2019]. Importantly, forward-mode AD is not applied to the portion $L(\boldsymbol{\epsilon})$ of the full loss $L_{\text{TD}}(\boldsymbol{\theta})$, which can involve costly simulations and AI processing. Instead, the gradient $\nabla L_{\boldsymbol{\epsilon}_0}$ required by GTRA is obtained via backward-mode AD, which is more efficient for scalar-valued functions. Thus, the computational cost of a GTRA iteration is roughly the cost of a conventional TRA iteration (mapping $\boldsymbol{\theta}$ to $\boldsymbol{\epsilon}$) plus the cost of forward and backward passes through the scalar loss (mapping $\boldsymbol{\epsilon}$ to $L(\boldsymbol{\epsilon})$), analogous to an SGD step. Since evaluating $L(\boldsymbol{\epsilon})$ can be much more expensive than ray tracing, GTRA provide an efficient practical solution.

Toy Task-Driven Design Problem. We revisit the toy problem of Sec. 3.2 in a task-driven setting, where we minimize $L_{\text{TD}}(\boldsymbol{\theta}) = {}^1/_2||\boldsymbol{I'}(\boldsymbol{\theta}) - \boldsymbol{I}||^2$: the scaled MSE between the virtual scene $\boldsymbol{I}$ and its simulated capture $\boldsymbol{I'}$, following the imaging simulation method described in Sec. 4. The lower portion of Fig. 4a shows the GTRA-optimized design using the converted residuals from Eq. (5). The optimization successfully retrieves an optimal solution for the chart scene made of concentric rings (PSNR: 24.8 dB), where the PSFs are narrow in the radial direction and wide in the angular direction, showing the flexibility of task-driven design. In contrast, the TRA-optimized design achieves a better effective spot radius, but its performance on this task is suboptimal, achieving only 20.0 dB.

From Fig. 4c and Table 1, as in the conventional design example, using the 1-dimensional residual

$$\boldsymbol{\ell}_{\text{MSE}}(\boldsymbol{\theta}) = [||\boldsymbol{I'}(\boldsymbol{\theta}) - \boldsymbol{I}||] \in \mathbb{R}^1 \tag{7}$$

is insufficient for convergence, even though it is equivalent to GTRA on a first-order basis, and SGD-based optimizers again fail to converge to a good solution. This demonstrates the effectiveness of the GTRA objective in a task-driven scenario.

Setup. We reproduce the setup of Sec. 3.2 exactly, except that $\boldsymbol{\ell}_{\text{TRA}}$ is replaced with $\boldsymbol{\ell}_{\text{GTRA}}$ in the effective approach, and $\boldsymbol{\ell}_{\text{ESR}}$ is

replaced with $\boldsymbol{\ell}_{\text{MSE}}$ in the naive approach. In imaging simulation, we consider a sensor of size 1024×1024 pixels and diagonal 11.55 mm. We consider a full-resolution chart as the virtual scene and apply a spatially varying convolution over 9×9 sensor regions. We do not consider diffraction effects in this particular problem.

3.4 Ray Tracing Overview

We describe our procedure for ray tracing, leaving details and validation against established software to Supp. S1.1 to S1.3.

Dispersion. We employ the Hartmann dispersion model [Lee 1926]

$$n(\lambda) = A + \frac{C}{\lambda - B}, \tag{8}$$

where $n(\lambda)$ is the refractive index at wavelength λ, and A, B, and C are analytically derived from glass variables.

Spot Diagrams. We obtain spot diagrams $\boldsymbol{\epsilon}$ by tracing rays from the object plane to the image plane, using the grating-modified law of refraction when modeling diffractive surfaces. In most experiments, we select $N_{\text{f}} = N_{\text{w}} = 11$ equidistant field points and wavelengths, and $N_{\text{p}} = 512$ pupil locations, resulting in 61 952 rays. We use precise iterative ray aiming [Côté et al. 2023b] to account for vignetting and pupil aberrations.

3.5 Optimization Overview

We next describe our optimization approach in its three core components: parameterization, optimization objective, and optimization algorithm. Additional details are provided in Supp. S2.1 to S2.3.

Parameterization. The optimizable variables include every surface curvature and spacing, except for the last curvature, solved to maintain either the focal length or image height. In some cases, the last spacing is also excluded—either solved to preserve the TTL or held fixed to maintain the WD. When aspheric surfaces are present, the even-asphere model describes the sag profile

$$\tilde{z}(r) = \frac{cr^2}{1 + \sqrt{1 - (1+\kappa)(cr)^2}} + \sum_{i=2}^{n+1} a_{2i} r^{2i} \tag{9}$$

as a function of the radial coordinate r, curvature c, conic constant κ, and aspheric coefficients a_{2i}. Diffractive surfaces are expressed by their phase profile

$$\phi(r) = \sum_{i=1}^{n} d_{2i} r^{2i}, \tag{10}$$

where d_{2i} are the diffractive surface coefficients. Rather than directly defining a_{2i} or d_{2i} as variables, we apply a linear transformation to express them in a more disentangled form, inspired by Q-type polynomials [Schuhmann 2019]. Glass variables are represented using at least 2 variables, the refractive index at 587.6 nm and the Abbe number. When normal partial dispersion cannot be assumed (e.g., for microscope objective lenses), we employ a 3rd variable, the deviation from normal partial dispersion [SCHOTT 2016]. Glass variables are normalized with a linear transformation based on the glass catalog employed, as in [Côté et al. 2021]. All aforementioned variables are scaled to be dimensionless.

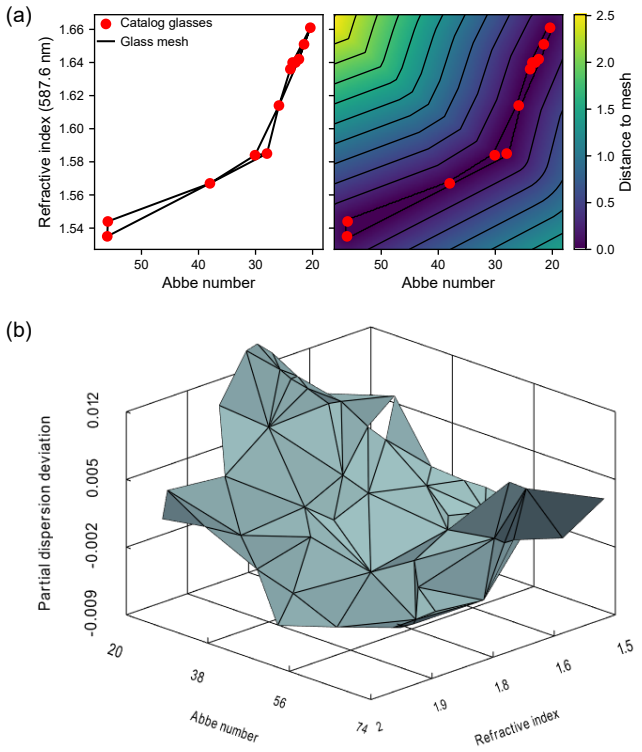

Fig. 5. **Glass catalog mesh for efficient glass optimization.** We present an automated method for constructing a mesh representation of available glasses, encouraging physically realizable solutions during optimization. The distance to the mesh serves as a penalty term included in the residuals. (a) For glass materials defined by refractive index and Abbe number, the catalog forms a 2D mesh, demonstrated here for the following smartphone telephoto lens study where 12 moldable (plastic) materials are available. (b) Including a third variable, such as the deviation from normal partial dispersion, creates a 3D mesh, shown here with 104 glass materials from the following microscope objective lens study.

Optimization Objective. The sum-of-squares objective $L(\boldsymbol{\theta}) = {}^1\!/_2 ||\boldsymbol{\ell}(\boldsymbol{\theta})||^2$ includes the TRA or GTRA residuals alongside other residuals that impose soft constraints on the design. We incorporate up to seven types of soft constraints, including *distortion residuals* and *relative illumination residuals* to control these image quality considerations within specified thresholds. Relative illumination residuals, when coupled with a *front aperture*—defined here as an auxiliary stop of fixed size and location in front of the first element—allow the optimizer to control the degree of vignetting, balancing relative illumination and aberration correction. In contrast, commercial optical software [Ansys, Inc. 2025; Keysight, Inc. 2025] typically keeps vignetting fixed.

The next three are geometry-related and re-use traced rays as probes to enforce well-behaved lens behavior, similar to Côté et al. [2023a]. *Ray path residuals* set minimum and maximum horizontal distances for rays between successive surfaces. *Ray angle residuals*

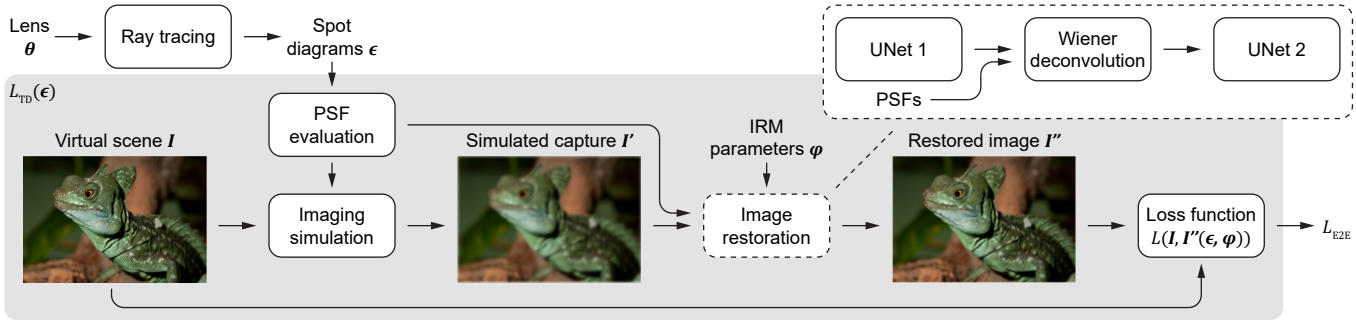

Fig. 6. **Task-driven loss function in E2E image restoration.** The E2E loss L_{E2E} quantifies the imaging system's ability to produce high-quality restored images. We simulate the image capture and restoration processes in a fully differentiable manner, aiming to minimize the difference between the restored image and the original scene. Our approach jointly trains an IRM that integrates data-driven techniques with classical methods. Importantly, the optics' task-driven loss $L_{\text{TD}}(\boldsymbol{\epsilon})$ is scalar-valued and depends solely on spot diagrams as a representation of the lens variables, fulfilling the conditions for GTRA.

ensure that the angle of incidence and refraction are within a tolerable range. *Surface normal residuals* prevent unfeasible aspheric element shapes by enforcing surface normals in a predefined range.

Finally, when glass materials are optimized, *glass mesh distance residuals* ensure that the optimized materials remain inside a 2D or 3D mesh generated to closely represent the volume of the glass catalog employed (Fig. 5). We couple them with *glass variable residuals*, with a weight typically augmented over time, to progressively minimize the distance between the optimized glass variables and their closest counterparts in the glass catalog.

Optimization Algorithm. The update rule for the LM algorithm is

$$\Delta\boldsymbol{\theta} = -\left(\mathbf{J}^T\mathbf{J} + \lambda\mathbf{D}^2\right)^{-1}\mathbf{J}^T\boldsymbol{\ell}_0\,, \tag{11}$$

where $\mathbf{D}^2$ is a diagonal matrix with positive diagonal entries that we designate as the *damping matrix*, and $\lambda \geq 0$ is the *damping factor*. To preserve scale invariance, we set the entries of $\mathbf{D}^2$ to the diagonal entries of $\mathbf{J}^T\mathbf{J}$ [Nocedal and Wright 2006]. We implement a rejection mechanism so that a given step is accepted if and only if $L(\boldsymbol{\theta} + \Delta\boldsymbol{\theta}) < L(\boldsymbol{\theta})$, which explains the monotonic decrease of LM curves in Fig. 4(b,c). This mechanism is essential as attempted steps can increase the loss by several orders of magnitude. We update the damping parameter λ using a standard heuristic [Transtrum and Sethna 2012]: if the loss decreases, λ is reduced (by a factor of 3) to favor Gauss-Newton behavior; otherwise, it is increased (by a factor of 2) to resemble gradient descent.

4 END-TO-END IMAGE RESTORATION

In this section, we outline our approach to solving the task of E2E image restoration, which aims to design an *imaging system*—including the lens, IRM, or other components—to maximize perceptual image quality. Central to this process is establishing a loss function that accurately reflects the system's ability to produce high-quality images (Fig. 6). The established approach [Metzler et al. 2020; Peng et al. 2019; Sitzmann et al. 2018; Sun et al. 2020; Yang et al. 2024a] involves modeling virtual scenes $\boldsymbol{I}$ with high-quality natural images, generating the simulated capture $\boldsymbol{I'}$, and processing the restored image $\boldsymbol{I''}$. The goal is to minimize the difference $L(\boldsymbol{I}, \boldsymbol{I''})$ between the restored image and original scene. We employ the MAE loss for IRM training steps, in line with standard practice, and MSE for lens optimization steps, for additional stability. We identify *image-driven design* as a special case of processing-aware design where the identity function is used for the IRM.

The imaging simulation is a crucial part of E2E optimization, as it must closely reflect reality to ensure the optimization is meaningful. We build on the simulation model from Côté et al. [2023a], which employs conventional ray-tracing techniques for optical design [Ansys, Inc. 2025; Keysight, Inc. 2025] to calculate spot diagrams, followed by geometric PSF evaluation and spatially varying convolution. Our improvements include an optimizable vignetting model and diffraction-compensated geometric PSFs.

4.1 PSF Estimation

Our PSF estimation method comprises a geometric stage based on kernel density estimation, followed by a diffraction-compensation heuristic. Additional details are provided in Supp. S1.4.

Geometric PSF Estimation. Spot diagrams, described in Sec. 3.4, serve as the primary input for our PSF estimation method. Rather than naively summing the rays, a non-differentiable operation, we use kernel density estimation (KDE) with a 2D linear kernel corresponding to twice the pixel size. We match the size of a virtual sensor bin to the physical sensor pitch, and use sufficient grid sizes to capture the PSF's spatial extent throughout optimization. The result of this operation is denoted $\boldsymbol{K} \in \mathbb{R}^{N_{\text{f}}N_{\text{w}}W_{\text{p}}H_{\text{p}}}$, where W_{p} and H_{p} are the width and height of the virtual detector.

Diffraction-Compensated PSFs. To account for diffraction effects—important in many optical systems—we introduce a heuristic. This models each ray as generating its own Airy disk pattern

$$\mathrm{U}_{\text{Airy}}(r) \propto \frac{2J_1\left(k\,\mathrm{NA}\,r\right)}{k\,\mathrm{NA}\,r}\,, \tag{12}$$

where J_1 is the Bessel function of the first kind (order 1), $k = {}^{2\pi}/_{\lambda}$ is the wavenumber, and r is the distance from the ray location at the image plane. All spanned Airy disk patterns are coherently summed to allow for both constructive and destructive interference. The coherent summation is key in closely replicating diffraction

PSFs obtained from industry-standard simulations. Our diffraction-compensated PSFs prevent the optimization process from favoring sub-diffraction-limited PSFs, which are physically unrealistic, over better solutions. While a rigorous diffraction treatment (e.g., using Huygens PSFs) would provide higher accuracy, our heuristic offers a favorable trade-off, capturing essential diffraction dynamics with geometric ray-tracing speed. We validate this heuristic against the diffraction PSFs of established software in Sec. 5.4.

4.2 Imaging Simulation

We simulate blur-inducing aberrations exclusively. Although distortion and relative illumination also contribute to image quality, they are better handled through soft constraints, as their direct simulation would inflate metrics like MSE in ways that overstate their perceptual significance. Supp. S1.5 provides additional details.

Spatially Varying Convolution. Our differentiable lens simulation model allows for a spatially varying, wavelength-dependent PSF $K(x, y, \lambda)$, as well as a different spectrum $S_C(\lambda)$ for each color channel C. Convolving the scene $I_C(x, y)$ with the PSF and adding noise $\eta(x', y')$ yields the simulated capture

$$I'_C(x', y') = \iiint K\left(x'-x, y'-y, \lambda\right)\, S_C\left(\lambda\right)\, I_C\left(x, y\right)\, d\lambda\, dx\, dy + \eta(x', y')\,. \tag{13}$$

In practice, all these quantities are discretized, and Eq. (13) corresponds to a spatially varying convolution operation.

Image restoration methods are often applied in a patchwise manner with moderate resolutions (e.g., 256×256). Over large sensor resolutions, such as those of modern smartphones, these patches are often small enough for the PSF to be considered spatially invariant. Regardless, we apply the spatially varying overlap-add method [Hirsch et al. 2010], and further divide each image patch into overlapping subpatches (typically 3×3), with each subpatch convolved with the PSF that best corresponds to this location. Such PSFs are obtained by averaging PSFs over the corresponding RGB color bands, which are weighted by the sensor spectral response $S_C(\lambda)$, then interpolating across field angles and rotating. The overlapping subpatches are then combined to obtain the final restored image, using a 2D Hann window to ensure smooth transitions.

Noise Model. We add Gaussian noise η with a standard deviation of 0.5 % of the maximum pixel value [Levin et al. 2009] to the simulated images, representative of well-lit capture conditions. While real sensors exhibit signal-dependent Poisson noise, we employ this simplified additive model to keep the restoration problem well-posed without introducing the complexity of full radiometric simulation.

4.3 Image Restoration

Our IRM architecture combines classical and deep learning methods (Fig. 6) by inserting Wiener deconvolution [Wiener 1949] between two distinct U-Net models [Ronneberger et al. 2015]. See Supp. S3.2 for further insights.

Implementation Details. In our single-stage conditioned IRM, both U-Net models follow an identical NAFNet architecture [Chen et al. 2022] with distinct weights. Each has a channel width of 16 and 12 blocks distributed across 4 resolution levels: 7 encoder blocks (1+1+1+4), 1 middle block, and 4 decoder blocks (1+1+1+1). The Wiener deconvolution includes a single learnable parameter representing the logarithm of the estimated SNR, initialized to 1000. In total, the architecture contains 3M parameters.

4.4 End-to-End Training

In our E2E approach, we alternate optimization between the lens and IRM, using MSE/GTRA with the LM optimizer for the lens and MAE with SGD-based optimizers for the IRM. See Supp. S3.3 for additional details.

E2E Training. Whether optimizing lenses, training IRMs, or performing joint optimization, we employ a patchwise training strategy. From the image dataset, we randomly crop patches of size 256×256 and form batches of 16. A sufficiently large batch size stabilizes the E2E lens optimization by ensuring each batch is representative of the dataset. In addition, since each image patch is attributed to a sensor region, we use a pseudo-random strategy to ensure well-distributed coverage of sensor regions in each batch. When employing image restoration, we train the IRM for 25k steps using the Adam optimizer [Kingma and Ba 2015] with an initial learning rate of 0.001 decayed to 0 via cosine annealing. In joint optimization, each IRM training step is interleaved with a lens update step. For image-driven lens optimization (without restoration), we likewise consider 25k optimization steps.

5 EXPERIMENTS

The following sections validate our general optics optimization approach on an array of diverse design tasks. Design files for all optimized lenses are included in the supplementary materials.

Secs. 5.1 and 5.2 validate our E2E method in two broadly deployed real-world applications: smartphone telephoto lenses and microscope objective lenses. These sections also address challenges such as material selection, vignetting control, and hybrid modeling with aspheric and diffractive surfaces. Supp. S4 and S5 provide additional details and insight for these design domains, respectively. Supp. S6 presents C-mount lenses as an additional application, comparing three strategies: conventional design, image-driven design without restoration, and full E2E design.

Sec. 5.3 focuses on *conventional* design, which serves as a broadly applicable testbench for comparing task-driven optimization strategies. Mathematically, conventional optics optimization is a special (and simpler) case of task-driven design. We hypothesize that optimization methods effective for conventional design will generalize to task-driven tasks, whereas ineffective ones will remain so.

In Sec. 5.4, we validate our diffraction-compensated geometric PSFs against the rigorous diffraction PSFs produced by CODE V.

5.1 End-to-End Design: Smartphone Telephoto Lenses

Smartphone telephoto lenses, as complex lenses that aim to minimize footprint while maintaining high image quality and satisfying multiple design constraints, present a challenging real-world problem for comparing our E2E design pipeline against conventional design. Telephoto lenses are characterized by their telephoto ratio

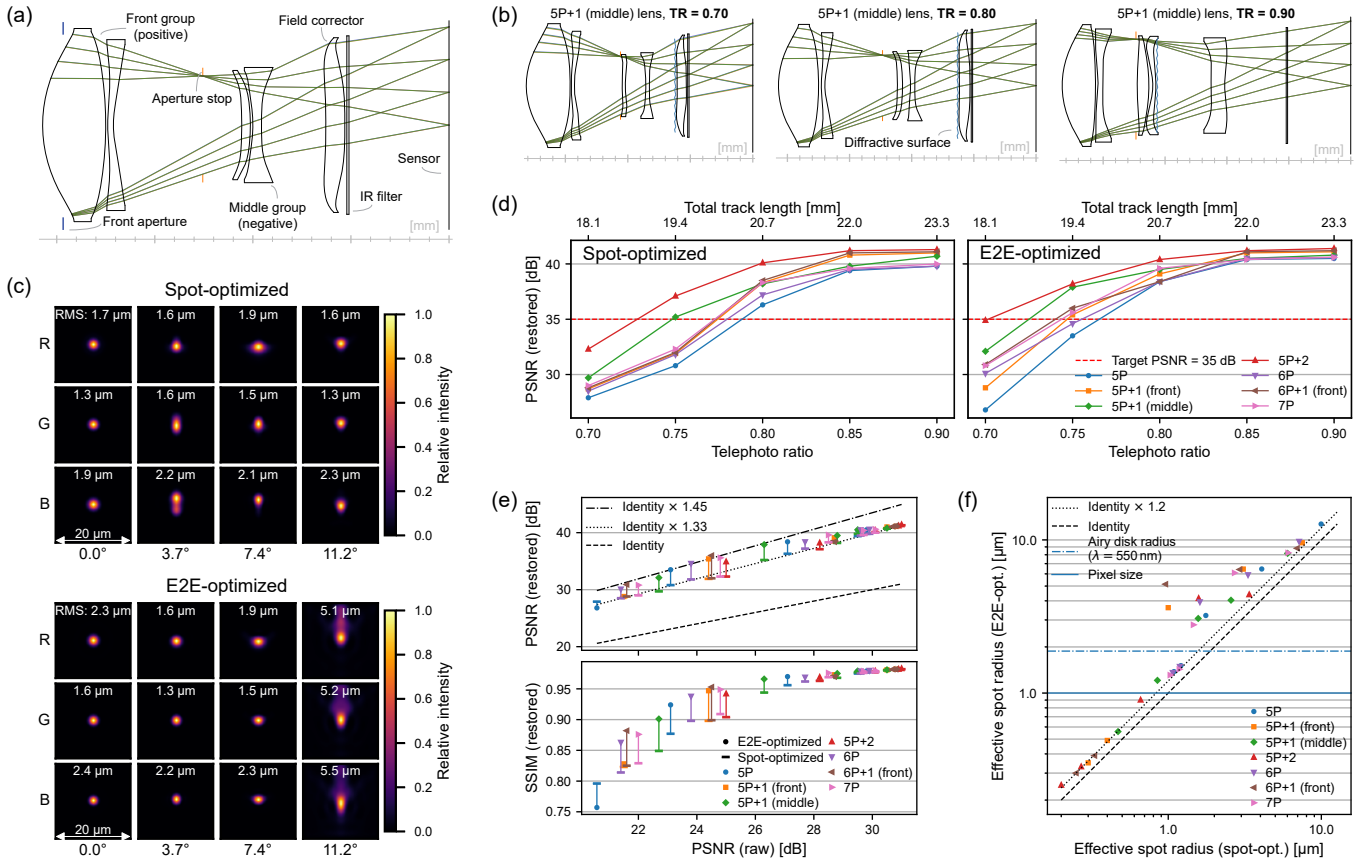

Fig. 7. **E2E design of smartphone telephoto lenses.** (a) Typical telephoto lens construction with 5 elements (denoted 5P) for a telephoto ratio (TR) of 0.8. (b) Hybrid lens configuration with a diffractive near the middle (5P+1), shown for different TRs. (c) PSFs of E2E-optimized lenses tend to have narrower, rounder central peaks but longer tails that explain their generally higher RMS sizes. (d) E2E optimization increases post-restoration PSNR (n=100 validation images), diminishing the TR requirement for meeting the image quality target. (e) Post-restoration PSNR and SSIM increase by consistent amounts based on pre-restoration PSNR, regardless of TR and configuration. (f) E2E-optimized lenses have a larger effective spot radius than their spot-optimized counterparts.

(TR), the ratio of TTL to effective focal length (EFL), with aberration correction becoming increasingly difficult as the ratio decreases [Smith 2004]. A typical smartphone telephoto construction is shown in Fig. 7a, and hybrid refractive-diffractive designs in Fig. 7b.

We study smartphone lenses with a magnification factor of 4.5×. We consider 7 different lens configurations, with 5–7 refractive elements and 0–2 ideal diffractive surfaces. For instance, the "5P" configuration includes 5 refractive elements (Fig. 7a), while the "5P+1 (middle)" configuration adds a diffractive surface next to the middle group (Fig. 7b). We vary each configuration across 5 TR values between 0.7–0.9 (Fig. 7b), where the TR is preserved by solving for the last airspace of the system, resulting in 35 combinations of configuration and TR. A total of up to 126 lens variables are optimized, depending on the configuration. See Supp. S4.1 for a detailed problem formulation.

Design Specifications. We consider a target EFL of 25.9 mm, a 22.31° FOV, and a target f-number of 2.8 to accommodate a 10.24 mm-diagonal sensor of resolution 8192×6144 with a 1 µm pixel pitch, inspired by the Samsung ISOCELL GN5. We use a typical smartphone sensor spectrum sampled at 11 wavelengths between 425–675 nm, weighted toward the center, which is grouped into RGB color bands for E2E optimization. We use 49×49-pixel PSF grids, corresponding to the sensor's 1 µm pitch.

Design Scenarios. To isolate the impact of E2E optimization from that of image restoration, each lens is optimized in two scenarios:

- *Spot-optimized lenses* are optimized for effective spot radius using the TRA objective. Although global optimality is rarely guaranteed, the following methodology strongly promotes high-quality solutions. We identify suboptimal solutions by comparing spot-radius trends: in lenses of the same configuration, spot radius should decrease monotonically with increasing TR; across configurations, we expect improved performance with more optical surfaces. If a solution appears

Table 3. **Minimum TR for satisfactory image quality in the telephoto lens study.** E2E-optimized lenses achieve the image quality target (PSNR ≥ 35 dB) with a shorter length than the corresponding spot-optimized lenses. See Fig. 7d for interpolated PSNR vs. TR curves. In hybrid configurations with a single diffractive surface (denoted as "+1"), its position within the lens system is indicated by "(front)" or "(middle)".

	Telephoto ratio (TR)		
Configuration	Spot-optimized	E2E-optimized	Improvement
5P	0.790	0.770	−0.020
6P	0.780	0.755	−0.025
5P+1 (front)	0.775	0.745	−0.030
7P	0.775	0.740	−0.035
6P+1 (front)	0.775	0.740	−0.035
5P+1 (middle)	0.750	0.725	−0.025
5P+2	0.730	0.700	−0.030

suboptimal, we restart optimization from a different initial design until results align with these expected trends. Each optimized lens is paired with a randomly initialized IRM instance, trained on the DIV2K dataset [Agustsson and Timofte 2017] to adapt to its specific optical response. Supp. S4.2 provides a detailed spot-radius assessment, while Supp. S4.3 offers an in-depth image quality evaluation with examples of both simulated and restored captures.

- *E2E-optimized lenses* are designed to minimize the MSE between virtual scenes and their restored images using the GTRA objective, as shown in Fig. 6. The corresponding spot-optimized lens is the starting point, and processing-aware optimization is interleaved with IRM training. In the E2E setting, we found that optimizing smartphone lens materials has a minimal impact on image quality and thus keep them fixed. *To demonstrate robustness, we optimize each of the 35 telephoto lenses in a single attempt using identical parameters.*

Design Constraints. In addition to TRA or GTRA residuals, we include soft constraints to the optimization objective for manufacturability and imaging properties, following industry standards where applicable. We enforce angles of incidence/refraction ≤60°, edge-to-thickness ratio between 1:3 and 3:1, distortion ≤2 %, airspaces ≥0.1 mm, element thickness between 0.25–4 mm, image clearance ≥5.1 mm, surface normals ≤30°, and relative illumination ≥70 %. With appropriately chosen weights, our optimization keeps residuals for soft constraints near zero, whether for spot-radius or E2E optimization. We therefore omit reporting final values for distortion, relative illumination, and similar constraints.

Image Quality Assessment. We evaluate the final PSNR and SSIM metrics on the DIV2K validation set (n=100 images) across all configurations, TRs, and design scenarios, with results summarized in Fig. 7. All validation images are centrally cropped to 256 × 256 pixels and mapped to distinct sensor regions.

Findings. Qualitatively, the E2E optimization process produces PSFs with narrower, rounder central peaks but longer tails (Fig. 7c). As the extended tails significantly inflate RMS values, such solutions would be discarded by conventional spot-radius optimization, highlighting a limitation of this criterion. In addition, E2E-optimized lenses show relatively uniform PSFs across the field, with the exception of corners where image quality is sacrificed. These findings hold across multiple configurations and TRs (see Supp. S4.4.2).

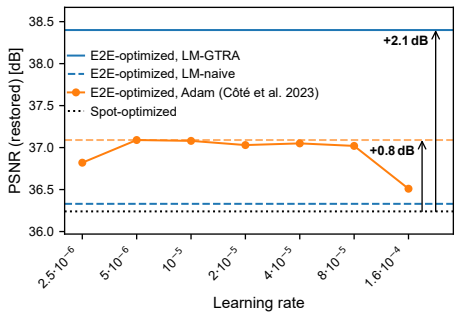

Fig. 8. **Comparison against first-order E2E optimization.** We compare our proposed LM-GTRA method against a first-order baseline adapted from [Côté et al. 2023a] using the Adam optimizer across a range of learning rates, applied to the 5P telephoto design depicted in Fig. 7a. Our method achieves a gain of 2.1 dB over the spot-optimized baseline (36.3 dB), significantly outperforming the best Adam result (0.8 dB). As shown in the LM-naive setting, LM performs poorly when using a mono-residual formulation for the scalar task-driven loss, highlighting the critical role of the GTRA objective.

By setting the threshold for post-restoration image quality to 35 dB—chosen so images qualitatively show minimal degradation, as shown in Supp. S4.3.2—we can interpolate the minimum TR for acceptable image quality for each configuration (Fig. 7d). TR improvements from E2E optimization range from 0.02–0.035 across all configurations (Table 3), resulting in a considerable 0.5–0.9 mm reduction in TTL (see Supp. S4.4.3 for validation with industry-standard software). For comparison, upgrading from the 5P to 7P configuration also yields an improvement within this range, as does adding an ideal diffractive surface to the front group.

E2E-optimized lenses consistently outperform spot-optimized lenses in post-restoration image quality (Fig. 7e). For spot-optimized lenses, the post-restoration PSNR improves over the pre-restoration PSNR by 33 %. For E2E-optimized lenses, this improvement rises to 45 % for moderate pre-restoration PSNR (23–26 dB), resulting in a net gain of about 3 dB. In contrast, high pre-restoration PSNR for spot-optimized lenses indicates that the PSFs are already diffraction-limited or subpixel, leaving little room for improvement. In addition, as by design, E2E-optimized lenses are outperformed by spot-optimized lenses in effective spot radius, typically by around 20 %, and in some cases up to 50–250 % (Fig. 7f).

Comparison with First-Order Baselines. To highlight the necessity of pseudo-second-order optimization in processing-aware design, we compare our GTRA-based LM method against a state-of-the-art first-order E2E baseline adapted from Côté et al. [2023a], which employs the Adam optimizer. We note that specific features of their method, such as the paraxial image solve and quantized-continuous glass variables, are not applicable here due to the TTL solve and fixed material selection. Similarly, the curriculum learning strategy of Yang et al. [2024a] is not applicable as our starting designs already

meet the target specifications. We apply the Adam baseline to the 5P telephoto design problem (TR=0.8) depicted in Fig. 7a, performing a grid search over multiple learning rates to ensure a fair comparison.

As shown in Fig. 8, our LM-based GTRA method achieves a final PSNR of 38.4 dB, a 2.1 dB improvement over the spot-optimized lens. In comparison, the best Adam-based result reaches 37.1 dB, an improvement of only 0.8 dB. This reinforces the findings from the doublet ablation in Sec. 3.2, demonstrating that the convergence advantages of our method scale to complex, high-dimensional real-world design problems. Notably, replacing our GTRA objective with a naive mono-residual formulation in the LM optimizer (LM-naive) yields poor performance, merely on par with the spot-optimized lens, confirming that the ray-level decomposition provided by GTRA is essential for unlocking LM's potential in processing-aware contexts.

Validation under Poisson-Gaussian Noise. To probe the sensitivity of our findings to the noise model of Sec. 4.2, we re-run the 5P design problem (TR=0.8) of Fig. 8 under a signal-dependent Poisson-Gaussian noise model with pixel-wise variance $\sigma^2 = \lambda_{\text{read}} + \lambda_{\text{shot}}\, x$ at noise-free pixel intensity $x \in [0, 1]$, fixing $\lambda_{\text{shot}} = 4.01 \times 10^{-4}$ and $\lambda_{\text{read}} = 3 \times 10^{-6}$ from Abdelhamed et al. [2018]. Both lenses are paired with IRMs retrained from random initialization under this model. The E2E-optimized lens reaches 35.0 dB PSNR against 33.7 dB for the spot-optimized baseline, a 1.3 dB margin. The absolute PSNRs are lower than under the Gaussian model as expected under this stronger noise. The image quality advantage of GTRA-based E2E optimization over spot-radius optimization thus persists under a more realistic noise regime.

5.2 Image-Driven Design: Microscope Objective Lenses

Microscope objective lenses, another class of complex real-world lens design challenges, can also benefit from processing-aware design. These small-FOV lenses pose distinct challenges compared to smartphone telephoto lenses, including larger apertures and greater emphasis on color correction across varied spectra, made possible by a much broader selection of glass materials. In this study, we apply our image-driven optimization approach, which minimizes the MSE between simulated and original images instead of relying on intermediate metrics. The aim is to preserve raw image quality for brightfield microscopy while increasing the WD of objective lenses, enabling a wider range of applications [Zhang and Gross 2019].

We study microscope objectives with 20× magnification and a numerical aperture of 0.4, following Côté et al. [2022]. We consider 3 lens configurations with 6, 8, or 10 lens elements (see Fig. 9a), involving up to 69 lens variables, including 3 for each material: refractive index at 587.6 nm, Abbe number, and deviation from normal partial dispersion. We vary each configuration across 5 WD values between 4–20 mm, covering normal, long, extra-long, and super-long WD regimes, resulting in 15 combinations of configuration and WD. See Supp. S5.1 for a detailed problem formulation.

We assume negligible tube lens aberrations and model the image formation process by simulating the objective lens in isolation and in reverse, treating the image and object as conjugate planes.

Design Specifications. Design specifications are based on the Atlas 16.8MP model featuring a 20.6 mm-diagonal Sony IMX387 CMOS sensor. Using 2×2 binning yields a 6.9 µm pixel pitch and a resolution of 2560 × 1536. We consider 81 × 81-pixel PSF grids with a pitch of 0.345 µm, matching the equivalent sensor pitch in object space. With a hypothetical 200 mm tube lens, we set the objective lens EFL to 10 mm, leading to an FOV of 5.9°. The spectrum is derived from the camera's quantum efficiency curves, sampled at 11 wavelengths between 450–650 nm.

Design Scenarios. As in the telephoto lens study, we consider two scenarios for each lens: spot-optimized (TRA) and MSE-optimized (GTRA) lenses, and consider sensible constraints in our optimization. No image restoration is applied in either case. We use the brightfield microscopy dataset BBBC041v1 [Ljosa et al. 2012].

Design Constraints. We enforce angles of incidence/refraction ≤60°, edge-to-thickness ratio between 1:3 and 3:1, distortion ≤2 %, airspaces ≥0.5 mm, and element thickness between 1–4 mm. In contrast to the telephoto study, we do not constrain surface normals since spherical surfaces are less prone to unfeasible shapes, nor relative illumination as these designs are vignetting-free.

Material Optimization. The challenge in glass material optimization is balancing sufficient exploration with the need to match final glass variables to catalog glasses. To efficiently explore the space of feasible materials, we represent our catalog of 104 Ohara-recommended glasses as a 3D mesh and penalize movement outside this volume. We include material optimization in both the conventional and processing-aware settings. During the first half of the optimization process (12.5k steps), we double the weight of the glass variable residuals every 500 steps and toggle glass variable optimization on or off every 250 steps, as strong constraints can slow convergence for other variables. At the midpoint of optimization, we bind the glass variables to the nearest catalog glass and exclude glass optimization for the remainder. We find that this strategy effectively optimizes and selects glass materials.

Image Quality Assessment. Since the BBBC041v1 dataset has a significant domain gap between training and validation sets, we first randomly repartition the dataset to ensure similar distributions, permanently designating 120 images for our validation set. We evaluate the final image quality metrics for all configurations, WDs, and design scenarios, following a methodology similar to that of the telephoto study, and compare the results in Fig. 9. Supp. S5.2 includes a spot-radius assessment, PSF and layout visualizations, and comparative analyses.

In addition, we perform a Monte-Carlo tolerancing analysis by generating random, uniformly distributed perturbations for each 6P lens (n=100 perturbed lenses) following industry-standard precision tolerances: ±0.01 mm for airspaces and element thicknesses, ±0.1 % for curvatures, ±0.0001 for refractive indices, and ±0.1 % for Abbe numbers. We refocus each lens as needed to maintain the defocus term and evaluate the average raw PSNR across all perturbed lenses.

Findings. MSE-optimized lenses outperform spot-optimized ones in raw PSNR in 14 of 15 settings (Fig. 9b), with improvements of 2.2, 2.0, and 1.6 dB for the 6P, 8P, and 10P lenses at WDs of 8, 16,

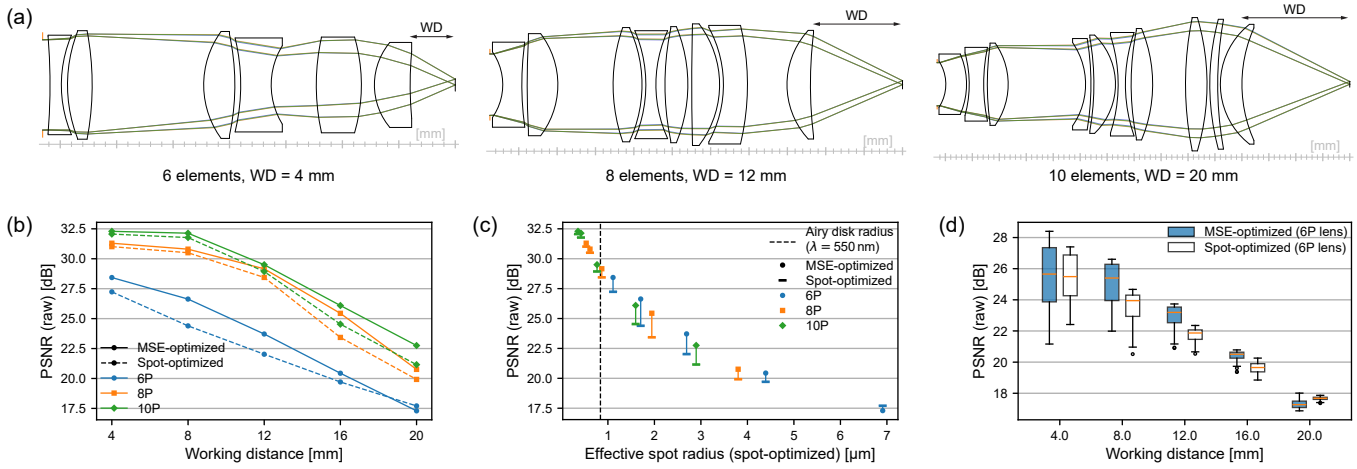

Fig. 9. **Image-driven design of microscope objective lenses.** (a) 2D layouts for different configurations (6P, 8P, 10P) and working distances (WDs). (b) MSE-optimized lenses outperform their counterparts in raw PSNR (n=120 validation images), resulting in increased WD for a given PSNR. (c) Improvements in raw PSNR from processing-aware optimization, as a function of baseline spot radius, are consistent across lens configurations and WDs. (d) Monte-Carlo tolerancing analysis for 6P lenses (n=100 perturbed lenses) showing that fabrication tolerances affect both spot- and MSE-optimized lenses similarly.

and 20 mm, respectively. The only outlier is the 6P lens at a WD of 20 mm, where its large PSF is not accurately captured due to the finite PSF grid size. At WDs of 12 mm and 16 mm, MSE-optimized 8P lenses surpass spot-optimized 10P lenses in raw PSNR. The PSNR improvement is lowest when the lenses are well below the diffraction limit, highest when geometric aberrations are dominant but reasonable, and moderate between these two regimes (Fig. 9c). Manufacturing imperfections, simulated via Monte-Carlo tolerancing, similarly affect both spot- and MSE-optimized lenses, preserving the average image quality gain from processing-aware optimization (Fig. 9d). While the lenses likely require compensators to reduce sensitivity to fabrication variability, this experiment supports that the GTRA optimization strategy does not exacerbate this sensitivity.

5.3 Conventional Design: Smartphone Wide-Angle Lenses

We replicate a design problem reported in Yang et al. [2024a], where a "curriculum learning" approach was used to generate lens designs from rudimentary starting points, without human intervention. This approach, relying on SGD-based optimization with expert-tuned learning rates, consists of progressively increasing the lens aperture and FOV. Our experiment builds on one of the reported designs (Fig. 10a): a smartphone wide-angle lens with 6 aspheric elements, a nominal focal length of 4.35 mm, an FOV of 82.8°, and an f-number of 1.98. Consistent with the original study, we evaluate design quality using spot diameter as the performance metric.

Replication Setup. We first produce a rough starting point (Fig. 10b) with all surfaces flat except for the last curvature, which is solved automatically to maintain the target focal length throughout optimization. Materials are set to the same dispersion curves as in Yang et al. [2024a]. Our design involves 84 of the original optimizable variables, including curvatures, spacings, and aspheric coefficients (degree 4 to degree 12) for each element excluding the solved curvature. Unlike Yang et al. [2024a], we omit degree-2 aspheric coefficients

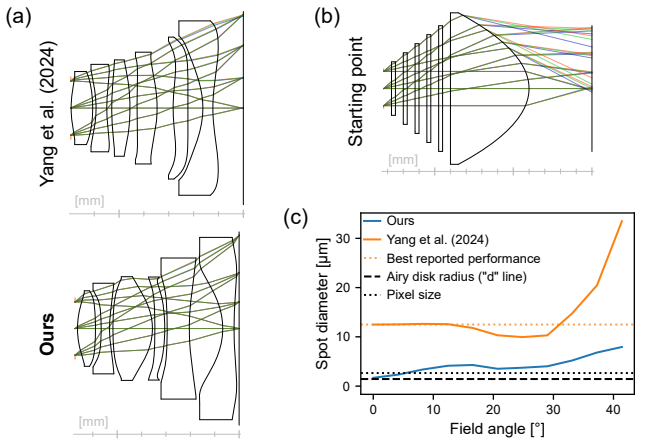

Fig. 10. **Conventional smartphone wide-angle lens design.** We replicate the conventional design experiment from Yang et al. [2024a]. (a) Layouts of the optimized designs from Yang et al. [2024a] and our method. (b) Rudimentary starting point for our autonomous optimization process, with glass materials fixed to the ones used in the original study. (c) Final spot diameter across the optical field, as evaluated in CODE V.

and fix conic constants to −1, as they are redundant with curvatures and degree-4 coefficients, respectively.

Optimization Objective. The optimization is guided by our LM method with the TRA objective. We consider 11 equally spaced field points and, following Yang et al. [2024a], wavelengths of 486, 587, and 656 nm. Additional constraints include ray path, ray angle, surface normals, and distortion residuals, with thresholds set to sensible values. Unlike Yang et al. [2024a], who allowed up to 15.9 % barrel distortion (evaluated with CODE V), we constrain distortion to within 2 %, reflecting stricter requirements.

Table 4. **Conventional optimization of smartphone wide-angle lenses from basic starting points.** We replicate the experiment of Yang et al. [2024a] to optimize a 6-element smartphone wide-angle lens from a rudimentary starting point, and compare to their reported design. We further fine-tune their design using the method of Côté et al. [2023a] for comparison. Our autonomously designed lens substantially outperforms the reference designs in spot diameter and distortion, as evaluated in CODE V.

	Avg. spot diameter [µm]	Distortion [%]
Yang et al. 2024 (reported lens)	14.7 (reported: 12.5)	15.9
Côté et al. 2023 (replicated method; best attempt is shown)		
From basic starting point	15.6	2.4
From Yang 2024 reported lens	8.4	2.7
Ours	**4.3**	**2.1**

Performance Evaluation. The result of our single automated optimization run is shown in Fig. 10c and summarized in Table 4. All lenses are evaluated in CODE V for consistency. Since CODE V does not support degree-2 aspheric coefficients, we refit curvatures and conic constants for the reference lens of Yang et al. [2024a] and re-optimize its last airspace using our method.

We also compare our method to Côté et al. [2023a], making a best-effort replication of their optimization approach, which applies the Adam optimizer with a paraxial image solve to keep the lens in focus throughout optimization. We perform a grid search over the global learning rate (10^{-3}, 10^{-4}, and 10^{-5}) and a scale factor (0.5, 1, and 2× the entrance pupil diameter) that effectively adjusts the learning rate relative to the dimensionality of each variable. We perform 20 000 Adam optimization steps with the same optimization objective and variables as used in our method. We consider both the basic design in Fig. 10b and the optimized design from Yang et al. [2024a] as starting points.

Findings. Our optimized lens achieves an average spot diameter of 4.3 µm, a *3.4× improvement* over the 14.7 µm result of Yang et al. [2024a]. This also exceeds their reported best result of 12.5 µm. Our method also outperforms the adapted approach from Côté et al. [2023a] by 3.6× when both methods start from the rudimentary lens, and by 1.9× even when their method is allowed to start from the final lens of Yang et al. [2024a]. These results underscore the limitations of first-order methods, as even the sophisticated curriculum learning approach of Yang et al. [2024a] produced a design that was significantly outperformed by our method. This experiment supports the need for effective optimization in optical design, which our GTRA optimization framework provides in the processing-aware context.

5.4 Validation of Diffraction-Compensated PSFs

In Fig. 11, we compare the PSFs from our framework to the diffraction PSFs of CODE V [Keysight, Inc. 2025], specifically for a 4-element telephoto smartphone lens depicted in Fig. 11a and further described in Supp. S1.3.4. This lens operates near the diffraction limit, with its geometric PSF size comparable to the Airy disk across most fields and wavelengths, making it an ideal candidate for evaluating diffraction effects.

We compare three PSF estimation settings:

- *Geometric PSFs* apply the KDE method without accounting for diffraction effects.
- *Naive Airy disk convolution PSFs* convolve the geometric PSF with the Airy disk intensity pattern directly, resulting in a broader spread than geometric PSFs as they cannot account for destructive interference.
- *Coherent Airy disk convolution PSFs* convolve the approximated field pattern of the geometric PSF with the Airy field pattern, then reconstruct the intensity (see Sec. 4.1).

To quantify the performance of the PSF estimation methods, we use the normalized root-mean-square error (NRMSE), where 0 % indicates a perfect match with the reference diffraction PSFs from CODE V and 100 % denotes a complete mismatch. Our results show that the coherent Airy disk convolution aligns both qualitatively and quantitatively most closely with the reference PSFs, achieving an average NRMSE of 4.2 % across all 11 fields and 9 wavelengths. In contrast, the naive Airy disk convolution permits only constructive interference, resulting in an overestimation of the PSF size in most cases, yielding an average NRMSE of 7.0 %. The geometric PSF underestimates the PSF size, as reflected in an average NRMSE of 15.5 %. These findings validate the effectiveness of our diffraction heuristic in capturing diffraction effects.

In Supp. S7.1, we further validate our PSF estimation for wide-angle systems by comparing against CODE V for the optimized wide-angle smartphone lens from Fig. 10a (bottom panel), which spans an 82.8° FOV. Our diffraction-compensated PSFs achieve an average NRMSE of 6.9 % across 11 fields and 9 wavelengths, confirming accurate PSF estimation even at large field angles where geometric aberrations dominate over diffraction.

6 CONCLUSION

We present a robust approach for processing-aware optical design based on *generalized aberrations*—a reformulation of the TRA objective suitable for scalar-valued task losses. By resolving the dimensionality mismatch between scalar task losses and high-dimensional optical parameters, this method unlocks pseudo-second-order optimization mechanisms that were previously restricted to conventional design. This framework effectively bridges the gap between traditional spot-radius optimization and modern E2E design, overcoming the convergence limitations of gradient-based optimization—an approach that, despite its widespread use in existing E2E works, frequently struggles with even low-complexity lens design. By highlighting and resolving these pitfalls, we aim to shift the E2E design paradigm toward these more robust, second-order solvers.

Our extensive simulations confirm the method's effectiveness across diverse design scenarios, including smartphone telephoto, microscope objective, and C-mount lenses. In evaluations of over 100 design instances, GTRA-based optimization consistently yields superior image quality compared to traditional spot-radius optimization and state-of-the-art first-order approaches. Crucially, this validates that our engine reliably optimizes what it is asked to optimize—a capability that was previously inconsistent in E2E frameworks. These

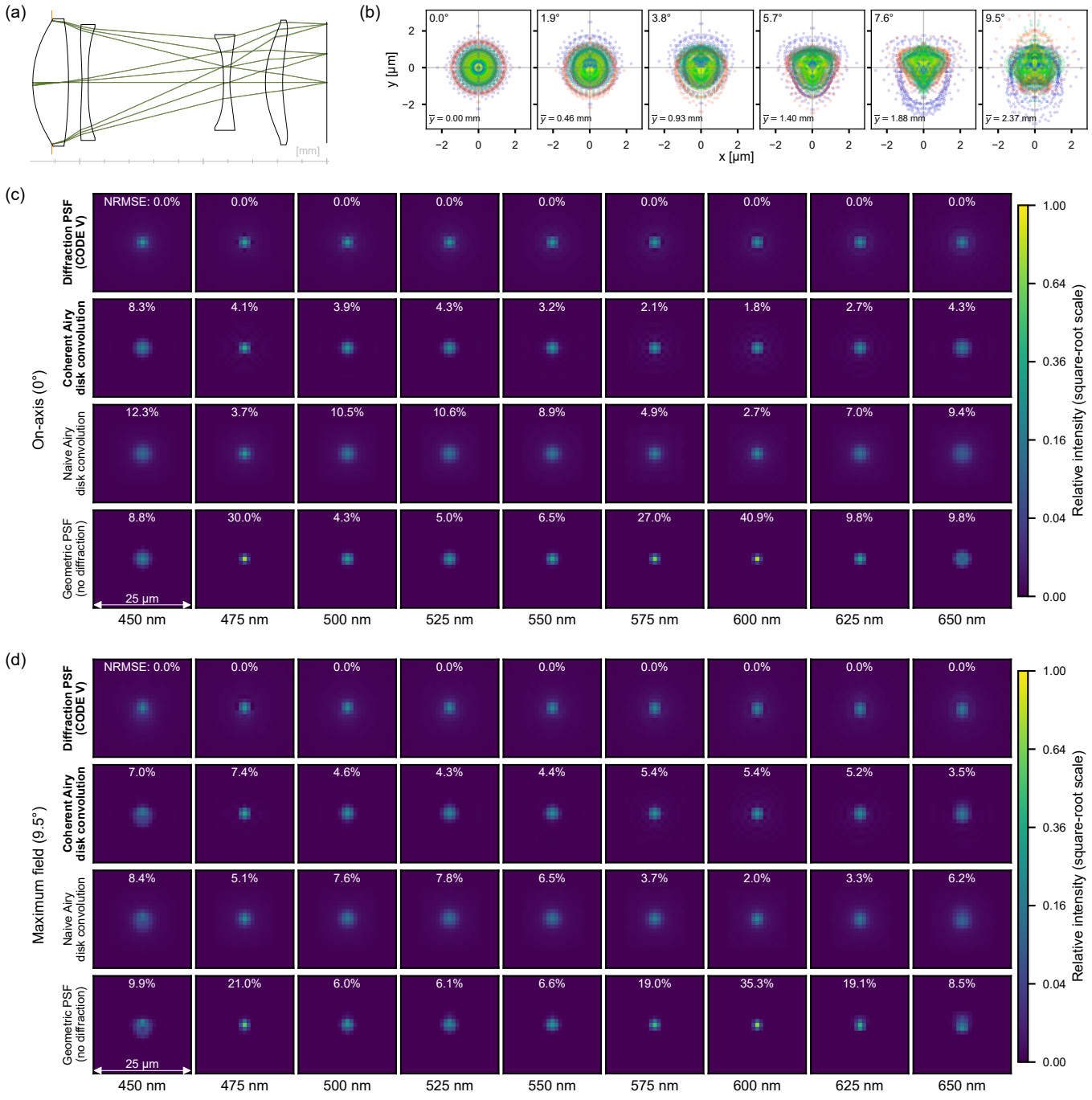

Fig. 11. **Comparison of PSF evaluation methods.** (a) 4-element smartphone telephoto lens used for this comparison. (b) Spot diagrams across different fields and wavelengths, serving as the input for the three considered PSF evaluation methods. (c) and (d) show the PSFs on-axis and at the maximum field, respectively, with the square root of the PSF shown to enhance the visibility of the tails. Our coherent Airy disk convolution qualitatively and quantitatively aligns most closely with the reference diffraction PSFs obtained using CODE V. In contrast, the naive Airy disk convolution allows only constructive interference, not destructive, leading to an overestimation of the PSF size in most cases. The geometric PSF, however, underestimates the PSF size.

improvements are partly explained by the optimizer's ability to exploit PSFs with narrow central peaks and elongated tails—features often penalized by conventional spot-radius metrics but beneficial for restoration.

While our experiments focus on image restoration, the GTRA formulation is inherently agnostic to the choice of task loss; extending it to downstream objectives such as object detection or semantic segmentation remains an open challenge. Incorporating rigorous differentiable diffraction handling would further improve physical accuracy, while more realistic sensor noise models would extend applicability to low-light and high-dynamic-range regimes. Physical fabrication and testing would, in turn, provide valuable insights into the remaining sim-to-real gap. We hope that, by demonstrating the pitfalls of first-order optimization in processing-aware design and offering a principled alternative, this work encourages the adoption of robust, second-order solvers across the broader landscape of E2E optical design.

ACKNOWLEDGMENTS

Ethan Tseng was supported by a Google PhD Fellowship. Ethan Tseng is a co-founder of Cephia AI. Felix Heide was supported by an NSF CAREER Award (2047359), a Packard Foundation Fellowship, a Sloan Research Fellowship, a Sony Young Faculty Award, a Project X Innovation Award, an Amazon Science Research Award, and a Bosch Research Award. Felix Heide is a co-founder of Algolux (now Torc Robotics), Head of AI at Torc Robotics, and a co-founder of Cephia AI.

REFERENCES

Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. 2016. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. *arXiv:1603.04467* (2016).

Abdelrahman Abdelhamed, Stephen Lin, and Michael S Brown. 2018. A high-quality denoising dataset for smartphone cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1692–1700.

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

Ansys, Inc. 2025. *Ansys Zemax OpticsStudio Documentation*. Ansys, Inc., Canonsburg, PA. https://support.zemax.com Version 2025 R2.

Gaurav Arya, William F Li, Charles Roques-Carmes, Marin Soljacic, Steven G Johnson, and Zin Lin. 2024. End-to-end optimization of metasurfaces for imaging with compressed sensing. *ACS Photonics* 11, 5 (2024), 2077–2087.

George Barbastathis, Aydogan Ozcan, and Guohai Situ. 2019. On the use of deep learning for computational imaging. *Optica* 6, 8 (Aug. 2019), 921–943. https://doi.org/10.1364/OPTICA.6.000921

Atilim Gunes Baydin, Barak A Pearlmutter, Alexey Andreyevich Radul, and Jeffrey Mark Siskind. 2018. Automatic differentiation in machine learning: a survey. *Journal of Marchine Learning Research* 18 (2018), 1–43.

Julie Bentley and Craig Olson. 2012. Field guide to lens design. Society of Photo-Optical Instrumentation Engineers (SPIE).

Praneeth Chakravarthula, Jipeng Sun, Xiao Li, Chenyang Lei, Gene Chou, Mario Bijelic, Johannes Froesch, Arka Majumdar, and Felix Heide. 2023. Thin on-sensor nanophotonic array cameras. *ACM Transactions on Graphics (TOG)* 42, 6 (2023), 1–18.

Julie Chang, Vincent Sitzmann, Xiong Dun, Wolfgang Heidrich, and Gordon Wetzstein. 2018. Hybrid optical-electronic convolutional neural networks with optimized diffractive optics for image classification. *Scientific reports* 8, 1 (2018), 1–10.

Julie Chang and Gordon Wetzstein. 2019. Deep Optics for Monocular Depth Estimation and 3D Object Detection. In *Proceedings of IEEE International Conference on Computer Vision*. 10192–10201. https://doi.org/10.1109/ICCV.2019.01029

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

Eunsue Choi, Gyeongtae Kim, Jooyeong Yun, Yujin Jeon, Junsuk Rho, and Seung-Hwan Baek. 2024. 360° structured light with learned metasurfaces. *Nature Photonics* 18, 8 (2024), 848–855.

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

Geoffroi Côté, Fahim Mannan, Simon Thibault, Jean-François Lalonde, and Felix Heide. 2023a. 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*. 20803–20812.

Geoffroi Côté, Simon Thibault, Jean-François Lalonde, and Felix Heide. 2023b. Joint optical design for computer vision tasks: challenges and solutions. In *International Optical Design Conference 2023*, Vol. 12798. SPIE, 51–55.

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

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

Johannes E Fröch, Praneeth Chakravarthula, Jipeng Sun, Ethan Tseng, Shane Colburn, Alan Zhan, Forrest Miller, Anna Wirth-Singh, Quentin AA Tanguy, Zheyi Han, et al. 2025. Beating spectral bandwidth limits for large aperture broadband nano-optics. *Nature Communications* 16, 1 (2025), 3025.

Yao Gao, Qi Jiang, Shaohua Gao, Lei Sun, Kailun Yang, and Kaiwei Wang. 2025. Exploring Quasi-Global Solutions to Compound Lens Based Computational Imaging Systems. *IEEE Transactions on Computational Imaging* (2025).

André Girard. 1958. Excerpt from Revue d'optique théorique et instrumentale. *Rev. Opt* 37 (1958), 225–241.

Herbert Gross. 2007. *Handbook of Optical Systems, Volume 3: Aberration Theory and Correction of Optical Systems*. Vol. 3.

Harel Haim, Shay Elmalem, Raja Giryes, Alex M. Bronstein, and Emanuel Marom. 2018. Depth Estimation From a Single Image Using Deep Learned Phase Coded Mask. *IEEE Transactions on Computational Imaging* 4, 3 (Sept. 2018), 298–310. https://doi.org/10.1109/TCI.2018.2849326

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

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

Philipp Holl, Vladlen Koltun, and Nils Thuerey. 2022. Scale-invariant learning by physics inversion. *Advances in Neural Information Processing Systems* 35 (2022), 5390–5403.

Wenzel Jakob, Sébastien Speierer, Nicolas Roussel, Merlin Nimier-David, Delio Vicini, Tizian Zeltner, Baptiste Nicolet, Miguel Crespo, Vincent Leroy, and Ziyi Zhang. 2022. *Mitsuba 3 renderer*. https://mitsuba-renderer.org.

Keysight, Inc. 2025. *CODE V Documentation*. Keysight, Inc., Santa Rosa, CA. https://www.keysight.com/nl/en/products/software/optical-solutions-software/optical-design-solutions/codev.html Version 2025.03.

Michael J Kidger. 2004. *Intermediate optical design*. Vol. 134. Spie Press, Bellingham, WA.

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

Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *Nature* 521, 7553 (May 2015), 436–444. https://doi.org/10.1038/nature14539

HW Lee. 1926. The Hartmann formula for the dispersion of glass. *Transactions of the Optical Society* 28, 3 (1926), 161.

Seokho Lee, Cherry Park, and Junsuk Rho. 2024. Mapping information and light: trends of AI-enabled metaphotonics. *Current Opinion in Solid State and Materials Science* 29 (2024), 101144.

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

Anat Levin, Yair Weiss, Fredo Durand, and William T Freeman. 2009. Understanding and evaluating blind deconvolution algorithms. In *2009 IEEE conference on computer vision and pattern recognition*. IEEE, 1964–1971.

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

Vebjorn Ljosa, Katherine L Sokolnicki, and Anne E Carpenter. 2012. Annotated high-throughput microscopy image sets for validation. *Nature methods* 9, 7 (2012), 637–637.

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

Joseph Meiron. 1965. Damped Least-Squares Method for Automatic Lens Design. *JOSA* 55, 9 (Sept. 1965), 1105–1109. https://doi.org/10.1364/JOSA.55.001105

Christopher A. Metzler, Hayato Ikoma, Yifan Peng, and Gordon Wetzstein. 2020. Deep Optics for Single-Shot High-Dynamic-Range Imaging. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, Seattle, WA, 1375–1385.

Jorge Nocedal and Stephen J Wright. 2006. Numerical Optimization 2nd edition.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. 2019. Pytorch: An imperative style, high-performance deep learning library. In *Advances in neural information processing systems*, Vol. 32. 8024–8035.

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

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. 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*. Springer, 234–241.

SCHOTT. 2016. *Refractive Index and Dispersion.* SCHOTT Technical Information TIE-29.

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

Vincent Sitzmann, Steven Diamond, Yifan Peng, Xiong Dun, Stephen Boyd, Wolfgang Heidrich, Felix Heide, and Gordon Wetzstein. 2018. End-to-end optimization of optics and image processing for achromatic extended depth of field and super-resolution imaging. *ACM Transactions on Graphics* 37, 4 (Aug. 2018), 1–13. https://doi.org/10.1145/3197517.3201333

Warren J. Smith. 2004. *Modern Lens Design.* McGraw Hill Professional, New York, NY.

David G Stork and M Dirk Robinson. 2008. Theoretical foundations for joint digital-optical analysis of electro-optical imaging systems. *Applied Optics* 47, 10 (2008), B64–B75.

Jipeng Sun, Kaixuan Wei, Thomas Eboli, Congli Wang, Cheng Zheng, Zhihao Zhou, Arka Majumdar, Wolfgang Heidrich, and Felix Heide. 2025. Collaborative on-sensor array cameras. *ACM Transactions on Graphics (TOG)* 44, 4 (2025), 1–18.

Qilin Sun, Ethan Tseng, Qiang Fu, Wolfgang Heidrich, and Felix Heide. 2020. Learning Rank-1 Diffractive Optics for Single-Shot High Dynamic Range Imaging. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, Seattle, WA, 1386–1396.

Qilin Sun, Congli Wang, Qiang Fu, Xiong Dun, and Wolfgang Heidrich. 2021. End-to-End Complex Lens Design with Differentiate Ray Tracing. *ACM Trans. Graph.* 40, 4, Article 71 (jul 2021), 13 pages. https://doi.org/10.1145/3450626.3459674

Mark K Transtrum and James P Sethna. 2012. Improvements to the Levenberg-Marquardt algorithm for nonlinear least-squares minimization. *arXiv e-prints* (2012), arXiv–1201.

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

Ethan Tseng, Ali Mosleh, Fahim Mannan, Karl St-Arnaud, Avinash Sharma, Yifan Peng, Alexander Braun, Derek Nowrouzezahrai, Jean-François Lalonde, and Felix Heide. 2021b. Differentiable Compound Optics and Processing Pipeline Optimization for End-to-End Camera Design. *ACM Trans. Graph.* 40, 2, Article 18 (jun 2021), 19 pages. https://doi.org/10.1145/3446791

M. van Turnhout and F. Bociort. 2009. Instabilities and fractal basins of attraction in optical system optimization. *Optics Express* 17, 1 (Jan. 2009), 314–328. https://doi.org/10.1364/OE.17.000314

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

Gordon Wetzstein, Aydogan Ozcan, Sylvain Gigan, Shanhui Fan, Dirk Englund, Marin Soljačić, Cornelia Denz, David AB Miller, and Demetri Psaltis. 2020. Inference in artificial intelligence with deep optics and photonics. *Nature* 588, 7836 (2020), 39–47.

Norbert Wiener. 1949. *Extrapolation, interpolation, and smoothing of stationary time series: with engineering applications.* The MIT press, Cambridge, MA.

C. G. Wynne. 1959. Lens Designing by Electronic Digital Computer: I. *Proceedings of the Physical Society* 73, 5 (May 1959), 777–787. https://doi.org/10.1088/0370-1328/73/5/310

Xinge Yang, Qiang Fu, and Wolfgang Heidrich. 2024a. Curriculum learning for ab initio deep learned refractive optics. *Nature communications* 15, 1 (2024), 6572.

Xinge Yang, Qiang Fu, Yunfeng Nie, and Wolfgang Heidrich. 2023. Image Quality Is Not All You Want: Task-Driven Lens Design for Image Classification. *arXiv e-prints* (2023), arXiv–2305.

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

Christopher Zach. 2014. Robust bundle adjustment revisited. In *European Conference on Computer Vision*. Springer, 772–787.

Qiangbo Zhang, Zeqing Yu, Xinyu Liu, Chang Wang, and Zhenrong Zheng. 2023b. End-to-end joint optimization of metasurface and image processing for compact snapshot hyperspectral imaging. *Optics Communications* 530 (2023), 129154.

Wenguan Zhang, Zheng Ren, Jingwen Zhou, Shiqi Chen, Huajun Feng, Qi Li, Zhihai Xu, and Yueting Chen. 2024. End-to-end automatic lens design with a differentiable diffraction model. *Optics Express* 32, 25 (2024), 44328–44345.

Yueqian Zhang and Herbert Gross. 2019. Systematic design of microscope objectives. Part I: System review and analysis. *Advanced Optical Technologies* 8, 5 (2019), 313–347.

Yuanlong Zhang, Xiaofei Song, Jiachen Xie, Jing Hu, Jiawei Chen, Xiang Li, Haiyu Zhang, Qiqun Zhou, Lekang Yuan, Chui Kong, et al. 2023a. Large depth-of-field ultra-compact microscope by progressive optimization and deep learning. *Nature Communications* 14, 1 (2023), 4118.