End-to-End Complex Lens Design with Differentiable Ray Tracing

QILIN SUN, King Abdullah University of Science and Technology, Saudi Arabia, Point Spread Technology, China
CONGLI WANG, King Abdullah University of Science and Technology, Saudi Arabia
QIANG FU, King Abdullah University of Science and Technology, Saudi Arabia
XIONG DUN, Point Spread Technology, China, Tongji University, China
WOLFGANG HEIDRICH, King Abdullah University of Science and Technology, Saudi Arabia

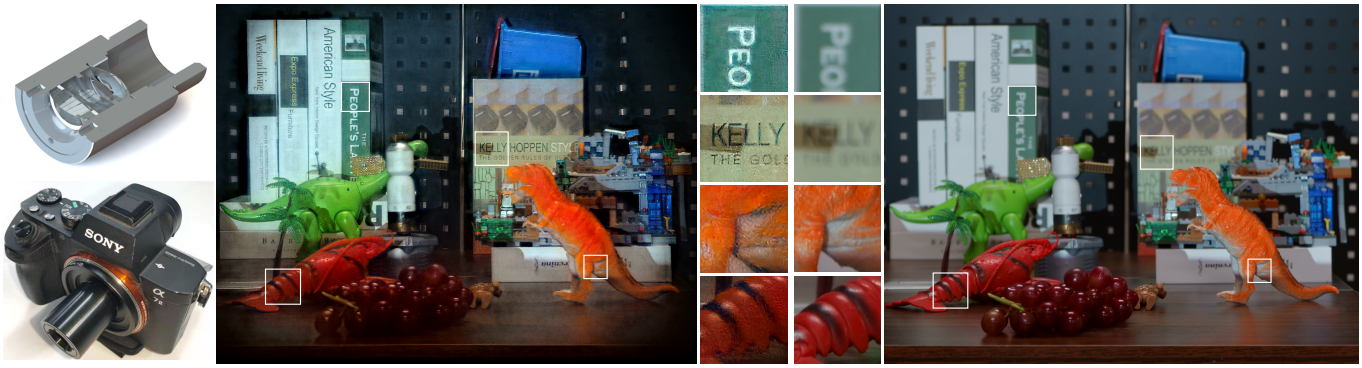

Fig. 1. An exemplary triplet design for extended depth-of-field imaging optimized by our end-to-end differentiable complex lens design framework. Top left: 3D model of the optimized triplet lens design ($50mm/F4$). Bottom left: prototype fabricated by single-point diamond turning. Middle: final image that is captured by our end-to-end designed lenses and processed by our algorithm. Right: the same scene captured by a Sony $28-70mm$ zoom lens at $50mm/F4.5$. The objects are placed in the range from around $0.8m$ to $1.8m$ from the lenses. Our prototype succeeds in obtaining the all-in-focus image, while the conventional lens shows a narrow depth of field.

Imaging systems have long been designed in separated steps: experience-driven optical design followed by sophisticated image processing. Although recent advances in computational imaging aim to bridge the gap in an end-to-end fashion, the image formation models used in these approaches have been quite simplistic, built either on simple wave optics models such as Fourier transform, or on similar paraxial models. Such models only support the optimization of a single lens surface, which limits the achievable image quality.

To overcome these challenges, we propose a general end-to-end complex lens design framework enabled by a differentiable ray tracing image formation model. Specifically, our model relies on the differentiable ray tracing rendering engine to render optical images in the full field by taking into account all on/off-axis aberrations governed by the theory of geometric optics. Our design pipeline can jointly optimize the lens module and the image reconstruction network for a specific imaging task. We demonstrate the effectiveness of the proposed method on two typical applications, including large field-of-view imaging and extended depth-of-field imaging. Both simulation and experimental results show superior image quality compared with conventional lens designs. Our framework offers a competitive alternative for the design of modern imaging systems.

CCS Concepts: • **Computing methodologies** → **Ray tracing**; **Computational photography**.

Additional Key Words and Phrases: Complex lens, Differentiable, Raytracing, End-to-end

ACM Reference Format:
Qilin Sun, Congli Wang, Qiang Fu, Xiong Dun, and Wolfgang Heidrich. 2021. End-to-End Complex Lens Design with Differentiable Ray Tracing. *ACM Trans. Graph.* 40, 4, Article 1 (August 2021), 13 pages. https://doi.org/10.1145/3450626.3459674

1 INTRODUCTION

Cameras are designed with a complicated tradeoff between image quality (e.g. sharpness, contrast, color fidelity), and practical considerations such as cost, form factor, and weight. High-quality imaging systems require a stack of multiple optical elements to combat aberrations of all kinds. At the heart of the design process are tools like ZEMAX and Code V, which rely on merit functions to trade off the shape of the PSF over different image regions, depth, or zoom settings. Such a design process requires significant user knowledge

Authors' addresses: Qilin Sun, Visual Computing Center, King Abdullah University of Science and Technology, Saudi Arabia,, Point Spread Technology, China, qilin.sun@kaust.edu.sa; Congli Wang, King Abdullah University of Science and Technology, Saudi Arabia, congli.wang@kaust.edu.sa; Qiang Fu, Visual Computing Center, King Abdullah University of Science and Technology, Saudi Arabia, qiang.fu@kaust.edu.sa; Xiong Dun, Point Spread Technology, China, Institute of Precision Optical Engineering, School of Physics Science and Engineering, Tongji University, China, dunx@tongji.edu.cn; Wolfgang Heidrich, Visual Computing Center, King Abdullah University of Science and Technology, Saudi Arabia, wolfgang.heidrich@kaust.edu.sa.

0730-0301/2021/8-ART1 $15.00
https://doi.org/10.1145/3450626.3459674

and experience, and the emphasis on PSF shaping neglects any subsequent image processing operations, specific application scenarios, or the desire to encode extra information in the image.

Therefore, domain-specific computational imaging has attracted researchers' attention in the past several decades. Enabling the co-design of optics with post-processing, computational cameras have achieved impressive results in extended depth of field (EDOF) [Cathey and Dowski 2002; Dowski and Cathey 1995; Levin et al. 2009; Tucker et al. 1999], high dynamic range imaging (HDR) [Debevec and Malik 1997; Mann and Picard 1994; Reinhard and Devlin 2005; Rouf et al. 2011], and image resolution [Brady et al. 2012; Cossairt et al. 2011; Nayar et al. 2004]. Nevertheless, all those older methods are either heuristic or use some proxy metric on the point spread function (PSF) rather than considering the imaging quality after post-processing. Therefore, finding a joint optimal solution for both imaging optics and image reconstruction for a given task remains an unsolved problem in general.

Over the past few years, co-design of optics and image processing [Peng et al. 2019; Sun et al. 2018], or even data-driven end-to-end design [Sitzmann et al. 2018] have emerged to bridge the gap between optical design and algorithm development. Co-design of optics and post-processing algorithms has achieved a superior performance for domain specific tasks such as depth estimation [Chang and Wetzstein 2019a], large field-of-view imaging [Peng et al. 2019], extended depth-of-field [Chang and Wetzstein 2019b], optimal sampling [Sun et al. 2020b], and high dynamic range (HDR) imaging [Metzler et al. 2020; Sun et al. 2020a]. Unfortunately, the differentiable lens models used in these works have been too limited to describe complex optical assemblies, and have instead only allowed to optimize a single optical surface with a single material. This narrow design space limits the final image quality compared to commercial consumer-level or industrial-level cameras. Furthermore, existing models are based on the paraxial approximation and ignore off-axis aberrations, which degrades the quality for large field of view imaging.

Data-driven optimization of all the parameters in a complex lens assembly is challenging. On the one hand, the varying parameters of the optical surfaces cause scaling and distortion that change during the optimization process. On the other hand, a naive implementation will consume huge computational resources due to the differentiable ray tracing engine [Nimier-David et al. 2019]. In this work, we overcome these challenges, and achieve the first end-to-end optimization system for complex lens assemblies. We propose a unique differentiable and configurable optical model that not only overcomes the limitation of a single optical surface and a single material, but also supports optimizing off-axis regions. In addition, we propose an end-to-end framework configurable for a given task with a tailored recovery algorithm, loss function, and data. As a result, we are able to directly render the images with aberrations of all kinds. That means we can optimize the complex lens model while accounting for the continuous variation of the PSF across the image plane. Beyond the goal of capturing a sharp and clear image on the sensor, the proposed method offers huge design flexibility that can not only find a compromise between optics and post-processing, but also opens up the design space for optical encoding.

It must be stressed, however, that our approach does not completely eliminate the need for an experienced user. Specifically, since lens design is a highly non-convex problem, we can not initialize the parameter space randomly; instead, we initialize the system with a coarse design that has the desired number of lens elements, and is roughly focused along the optical axis. This optical system is then improved and adapted to a specific imaging scenario using end-to-end optimization. In this paper we demonstrate both large field-of-view and large depth-of-field as the two application scenarios. The proposed approach outperforms the state-of-the-art complex lens design (by ZEMAX) in both simulation and experiments. We prototype our designs with complex lenses manufactured by a CNC machining system that supports point diamond turning. The experiments are carried out in-the-wild as conventional cameras. Our results show significantly improved performance over conventional designs on the above-mentioned applications.

Specifically, we make the following contributions:

- We introduce a novel configurable and differentiable complex lens model based on differentiable ray-tracing, and this model can simulate aberrations of all kinds. We allow users to easily define the initial optics design, including lens surface profile, positions, and materials.
- Our differentiable complex lens model is the first in end-to-end design to consider off-axis performance, and offers a greater design freedom compared to existing end-to-end optics models.
- We propose an end-to-end pipeline that can jointly optimize the lens model and the recovery network. The reconstruction network and loss functions can be tailored to a given computational imaging task.
- We successfully apply our model and pipeline to large field-of-view imaging and extended depth-of-field imaging using designs that are compact and low-budget, but high-quality. We validate them in both simulations and on real-world measurements captured by our assembled and fabricated aspherical lens group and verify that the experimental results agree with the simulations.

2 RELATED WORK

Optical Aberrations and Traditional Lens Design. The most common monochromatic aberrations are defocusing, spherical aberration, coma, astigmatism, field curvature, and distortion, while the chromatic aberrations are typically axial and lateral chromatic aberration. Both types of aberrations are the result of the differences in the optical path length when light travels through different regions of a lens at different incident angles [Fowles 2012]. These aberrations manifest themselves as an unwanted blur, which becomes more severe with increasing depth of field (DOF), numerical aperture, and field of view (FOV) [Smith 2005].

Conventional lens design is a semi-automated process, in which a rough initial design is chosen by an experienced designer, and then optimized with software like CODE V and ZEMAX. These typically use either the Levenberg-Marquardt algorithm or damped least squares (DLS) to optimize the optical system including spherical and aspherical lenses, hybrid optical elements [Flores et al. 2004;

Liu et al. 2007], and lens elements with different material properties. These tools are the cornerstone of lens design and rely on existing aberrations objectives, so-called merit functions, to find a compromise across a variety of criteria [Malacara-Hernández and Malacara-Hernández 2016; Shih et al. 2012], trading off the PSFs across sensor locations, lens configurations (e.g., zoom levels), and target wavelength band.

However, critically the established merit functions only operate on the PSFs, trading off their footprint over different configurations. This approach is agnostic to any intended usage case or image reconstruction approach. As a result, it is hard to co-design the optics and post-processing together for domain-specific cameras [Sitzmann et al. 2018] since they can not use the final imaging performance criteria as an optimization object. Thinking beyond the traditional complex lens design for a given task, we seek to investigate a differentiable complex lens model and end-to-end optimization framework to bring the complex lens design into an end-to-end era.

Computational Optics. Many works on computational imaging [Levin 2010; Levin et al. 2007; Stork and Gill 2013, 2014] have proposed designing optics for aberration removal in post-processing. These methods often favor diffractive optical elements (DOEs) [Antipa et al. 2018; Dun et al. 2020; Heide et al. 2016; Monjur et al. 2015], or even metasurfaces [Colburn et al. 2018] over refractive optics because of their large design space or ultra thin form factor [Khan et al. 2019]. To simplify the inverse problem in post-processing, all of the described approaches ignore off-axis aberrations by restricting the FOV to a few degrees – existing methods do not realize monocular and chromatic imaging with a large FOV. The state-of-the-art joint designing of optics and post-processing [Peng et al. 2019] firstly enables a large FOV imaging with a single lens. However, their model is still to design the optics and image processing algorithm separately to include the FOV in the design process. Moreover, they need a complicated and time-consuming dataset acquisition from the monitor.

In addition to minimizing optical aberrations optics, computational imaging also aims to improve the basic capabilities of a camera by including optical coding, such as depth of field [Cathey and Dowski 2002; Dowski and Cathey 1995; Levin et al. 2009; Tucker et al. 1999], dynamic range [Debevec and Malik 1997; Mann and Picard 1994; Reinhard and Devlin 2005; Rouf et al. 2011] and image resolution [Brady et al. 2012; Cossairt et al. 2011; Nayar et al. 2004].

Our proposed end-to-end complex lens design framework could be applied to many of these applications. It introduces a general design paradigm for computational cameras that optimizes directly for the post-processed output with respect to a chosen quality metric and domain-specific dataset.

End-to-end Optics Design. Co-designing of optics and post-processing has demonstrated superior performance over traditional heuristic approaches in single-lens color imaging [Chakrabarti 2016; Peng et al. 2019], HDR imaging [Metzler et al. 2020; Sun et al. 2020a], single image depth estimation [Boominathan et al. 2020; Chang and Wetzstein 2019a; Haim et al. 2018; Kotwal et al. 2020; Wu et al. 2019a,b; Wu et al. 2020; Zhang et al. 2018], microscopy imaging [Horstmeyer et al. 2017; Kellman et al. 2019; Nehme et al. 2019; Shechtman et al. 2016].

In computer vision, the emergence of deep learning has led to rapid progress in several challenging tasks and the state-of-the-art results for well-established problems [Schuler et al. 2013; Xu et al. 2014; Zhang et al. 2017]. For example, a deep approach for deconvolution by including a fully connected convolutional network [Nah et al. 2017] has been proposed. Generative adversarial networks (GANs) are shown to provide generative estimates with high image quality. Kupyn *et al.* [2017; 2019] demonstrated the practicability of applying GAN reconstruction methods to deblurring problems. Those approaches have been demonstrated to obtain state-of-the-art results in many computational photography tasks but not take one step further to optimize the optics together. G. Côté *et al.* [2019; 2021] utilize deep learning to get lens design databases to produce high-quality starting points from various optical specifications. However, they generally focused on the design of starting points, and the designing space is limited to spherical surfaces.

The deep optics [Sitzmann et al. 2018] approach involves joint design of optics and image recovery for a specific task in an end-to-end fashion. Based on this model, a series of applications have been investigated in the last two years like hyperspectral imaging [Baek et al. 2020; Jeon et al. 2019], high dynamic range imaging [Metzler et al. 2020], full-spectrum imaging [Dun et al. 2020] and depth estimation [Chang and Wetzstein 2019b]. These works are inherently limited to designing only a single optical surface, and therefore the image quality of their final designs does not reach the level of regular consumer camera optics. A solution to this problem has been to utilize a commercial lens but add a single, co-designed element for a specific purpose. This approach has been applied to super-resolution SPAD cameras [Sun et al. 2020b] and high dynamic range imaging [Sun et al. 2020a]. However, none of these approaches can deal with large FOVs as their image formation model relies on simple paraxial approximation in addition to the single-surface restriction. Co-designing complex optics with the image reconstruction was not addressed until the work on learned large FOV imaging [Peng et al. 2019]. They overcame the limitation of FOV by separating the optical design and image processing but not in an end-to-end fashion.

In conclusion, existing end-to-end methods work in a very restricted setting, including only a single optical surface and a simple paraxial image formation model (small FOV), or rely on existing optical design tools. They also have in common that they require either accurate PSF calibration or extensive training data. We propose a general configurable and differentiable complex lens model and an end-to-end framework with tailored recovery networks for different tasks. Drawing inspiration from the state-of-the-art differentiable rendering technique [Bangaru et al. 2020; Nimier-David et al. 2019; Zhang et al. 2020, 2019], our complex lens model offers a great design freedom where the number of elements, lens surface profiles and positions can be configurable. The ample design space of our proposed lens model allows for rich optical encodings and the end-to-end pipeline achieves optimal synergy with the image reconstruction algorithm. Our complex lens model can optimize the lens parameters and simulate all kinds of aberrations without considering spatial and depth varying PSF. This property makes it easier for the later reconstruction network retraining and fine-tuning stage to get a highly accurate simulated dataset. Finally, our

solution overcomes the limitation for large FOV and makes it possible to design a high-quality consumer-level lens in an end-to-end manner.

3 END-TO-END OPTIMIZATION OF COMPLEX LENS AND IMAGE RECOVERY

3.1 Image Formation Model

Our end-to-end framework consists of an optical simulation stage with the lens model and a trimmed recovery network as a reconstruction stage to achieve the best results by employing a generative adversarial network (GAN). As in most existing complex lens systems, the refraction is usually generated by either spherical or aspherical surfaces. Throughout the rest of this paper, we consider rotationally symmetric lens profile designs, which can be manufactured using diamond turning machines.

Note, however, that our lens model could be easily applied to rotationally asymmetric profiles such as Zernike basis functions.

3.1.1 Differentiable Lens Model and Ray Tracer. We implement our differentiable lens tracer following [Kolb et al. 1995], based on Mitsuba2 [Nimier-David et al. 2019]. The framework is fully differentiable, including the configurable and differentiable complex lens model and the image recovery network. Each part of this pipeline can be easily configured to tailor for specific tasks.

After training, we take the optimized parameters like radius, conic coefficients and high order coefficients of the lens profiles to fabricate the lens. To account for better reconstruction with the image processing pipeline and the recovery network, it can be tailored and fine-tuned through re-training after the lens parameters are fixed.

In the following, we show how to efficiently integrate the differentiable ray-tracing into our lens designing pipeline.

Aspherical Lenses. Our lens model is based on a standard representation of an aspherical lens as a spherical component with a polynomial correction factor. Given a Cartesian coordinate system (x, y, z), the z-axis coincides with the optical axis, while (x, y) forms the transverse plane. Let $r = \sqrt{x^2 + y^2}$ and $\rho = r^2$. Then the height of the aspheric surface and its derivative is defined as:

$$h(\rho) = \frac{c\rho}{1 + \sqrt{1 - \alpha\rho}} + \sum_{i=2}^{n} a_{2i}\rho^i, \tag{1}$$

$$h'(\rho) = c\frac{1 + \sqrt{1 - \alpha\rho} - \alpha\rho/2}{\sqrt{1 - \alpha\rho}\left(1 + \sqrt{1 - \alpha\rho}\right)^2} + \sum_{i=2}^{n} a_{2i} i \rho^{i-1}, \tag{2}$$

where c is the curvature, $\alpha = (1 + \kappa)c^2$ with κ being the conic coefficient, and a_{2i}'s are higher-order coefficients. The implicit form $f(x, y, z)$ and its spatial derivatives ∇f are:

$$f(x, y, z) = h(\rho) - z, \tag{3}$$

$$\nabla f = \left(2h'(\rho)x, 2h'(\rho)y, -1\right). \tag{4}$$

Note that spherical surfaces are special cases of aspheric surfaces when $\kappa = 0$ and $a_{2i} = 0\,(i = 2, \cdots, n)$.

In the following, we derive a differentiable ray-tracing based image formation model which simulates all kinds of aberration at the same time. For each surface in the lens, its profile is directly described by (1) and the lens materials are predefined according to the prior knowledge of optical design to cancel the chromatic aberrations.

Ray-surface Intersection by Newton's Method. To use the above lens model in a ray-tracer, we need to be able to compute the intersection point (x, y, z) and ray marching distance t for intersecting surface $f(x, y, z) = 0$ (implicit form), given a ray $(\mathbf{o}, \mathbf{d})$ of origin $\mathbf{o} = (o_x, o_y, o_z)$ and direction $\mathbf{d} = (d_x, d_y, d_z)$ of unit length (i.e. $\|\mathbf{d}\| = 1$). Mathematically, this is a root finding problem, i.e. we need to determine $t > 0$ such that

$$f(x, y, z) = f(\mathbf{o} + t\mathbf{d}) = 0. \tag{5}$$

Since there is no analytical solution for this problem for the aspherical lens model, we solve the problem numerically using Newton's method. At iteration $k+1$, we update $t^{(k+1)}$ from previous estimate $t^{(k)}$ as:

$$\begin{aligned} t^{(k+1)} &\leftarrow t^{(k)} - \frac{f(\mathbf{o} + t^{(k)}\mathbf{d})}{f'(\mathbf{o} + t^{(k)}\mathbf{d})} \\ &\leftarrow t^{(k)} - \frac{f(\mathbf{o} + t^{(k)}\mathbf{d})}{\nabla f \cdot \mathbf{d}}, \end{aligned} \tag{6}$$

where f' and ∇f denote derivatives w.r.t. t and (x, y, z), respectively. A coarse (non-singular) initialization is $t^{(0)} = (z - o_z)/d_z$, and the iteration stops when the difference is smaller than tolerance.

Dispersion by Cauchy's equation. To model dispersion, we extend Mitsuba2 by formulating the lens material refractive index using Cauchy's equation [Jenkins and White 2018]:

$$n(\lambda) = A + \frac{B}{\lambda^2} + \frac{C}{\lambda^4} + \cdots. \tag{7}$$

In practice, we found it sufficient to use only the first two terms (with parameters A and B) in the equation. When the central wavelength n_D and Abbe numbers V are given, A and B are computed as:

$$A = n_D - \frac{B}{\lambda_D^2} \qquad \text{and} \qquad B = \frac{n_D - 1}{V(\lambda_F^{-2} - \lambda_C^{-2})}, \tag{8}$$

where $\lambda_D = 589.3\,\text{nm}$, $\lambda_F = 486.1\,\text{nm}$, and $\lambda_C = 656.3\,\text{nm}$.

3.1.2 Optics simulation. End-to-end computational imaging consists of simulated optics used to generate simulated image data with all aberrations present, as well as software reconstruction pipeline. For joint design, both of these modules should be fully differentiable so that gradient update become possible across both components.

With the optimal trade-off between the simulation stage and the reconstruction stage, the PSF usually varies within the field of view, and across scene depth and spectrum. For a given color channel c, the recorded sensor measurement I_c can be expressed as:

$$I_c(x', y') = \int Q_c(\lambda) \cdot [p(x', y', d, \lambda) * s_c(x', y', d)]d\lambda + n(x', y'), \tag{9}$$

where the PSF $p(x', y', d, \lambda)$ is a function with spatial position (x', y') on the sensor, the depth d of scene, and the incident spectral distribution λ. Q_c is a function of the color response of the sensor, and $s_c(x', y', d)$ and $n(x', y')$ represent the latent scene and measurement noise (white Gaussian noise), respectively. The operator $*$ represents convolution.

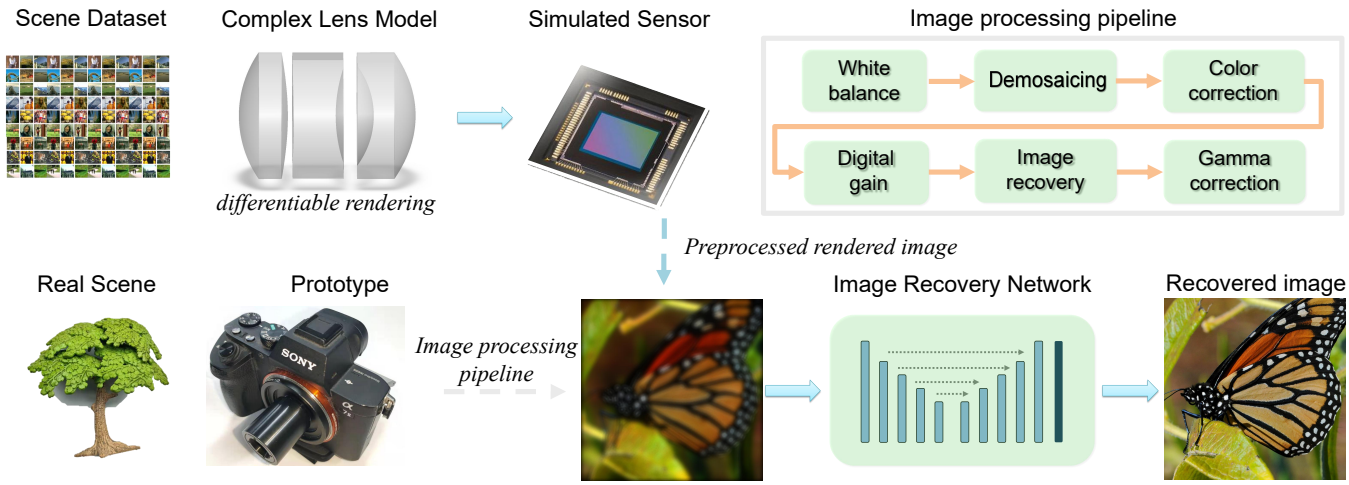

Fig. 2. Framework for end-to-end designing of differentiable complex lens model and reconstruction. In each forward pass, we set up one scene from a certain point-of-view and render the simulated sensor image through the differentiable complex lens model. Then, the simulated images are sent to the image reconstruction network and we train the whole framework simultaneously. For the experimental stage, we directly send the preprocessed real-captures to the pre-trained network. Notice that the scene setup, initial lens design and image recovery network can be tailored to specific applications.

We use Monte Carlo sampling in the rendering engine. At each pixel, rays are sampled starting from the sensor plane, with the wavelengths, sub-pixel origin shift, and direction sampled by a uniform random number generator without any importance sampling. These sampled rays are then traced sequentially through each of the refractive surfaces following Snell's law. Rays are marked as invalid and do not contribute to the final rendered image when the intersections are outside of the lens geometry or when total internal reflection takes place.

Unfortunately, the number of samples per pixel (SPP) is limited by the GPU memory, resulting in Monte Carlo rendering noise. To overcome this issue, we first render several passes and average them to get a clean estimate, then replace the PyTorch variable with the clean estimation to calculate the gradient multiple times to get the averaged clean gradients. After this processing, the obtained images and gradients are clean enough, and the Monte Carlo sampling noise can be ignored [Guo et al. 2018] compared to added white Gaussian noise.

3.1.3 Image alignment during training. Another challenge for end-to-end optical design is that in the initial stages of the optimization, the simulated image is both distorted and scaled compared to the desired reference. This misalignment makes it hard to accurately calculate a meaningful loss between the reference image and the rendered simulations. To solve the problem of pixelwise alignment, we first forward trace 16 points that are uniformly distributed from the center of the texture plane to the border and obtain the points $\mathbf{r}_d$ intersected with the sensor plane. Then we set corresponding ideal points on the sensor as $\mathbf{r}$ and the relation between the point pairs can be expressed as:

$$\mathbf{r}_d = \xi \mathbf{r}(1 + k_1\mathbf{r}^2 + k_2\mathbf{r}^4 + k_3\mathbf{r}^6) \tag{10}$$

To simulate the distortion and magnification, we only consider the radial distortion and solve a least-squares problem as:

$$\min_{\mathbf{K}} \|[\mathbf{r}, \mathbf{r}^3, \mathbf{r}^5, \mathbf{r}^7]\mathbf{K}^T - \mathbf{r}_d\|_2^2, \tag{11}$$

where $\mathbf{K} = \xi(1, k_1, k_2, k3)$ represents the current distortion coefficients along with a magnification coefficient ξ. Then we distort and resize the reference ground truth to match the currently rendered simulation pixel-to-pixel. Once the lens parameters are fixed, we undistort the captured image in the experiments.

3.2 Image Reconstruction

End-to-end lens design. As shown in Figure 2, we connect a U-net like architecture [Chen et al. 2018] with deep layers trimmed (only use the marked layers in Figure 3 in designing stage) but its early layers filters that can encode the information on sensor [Peng et al. 2019]. This setup speeds up the training process and provides sufficient degrees of freedom to encode the simulated information for the end-to-end design. Specifically, the trimmed U-net architecture in the design stage has three scales with two max pool operations for downsampling and two transposed convolutions for upsampling. At the bottleneck, we adopt two flat convolutional layers. Each convolutional layer is followed by a parametric rectified linear unit (PRelu). The trained weights in this trimmed U-net network are then taken to initialize the corresponding layers for the final fine reconstruction. Refer Figure 3 for details.

Generator. At the final reconstruction stage with fixed lens parameters, we adopt a GAN as shown in Figure 3 to recover the corrupted sensor image I from the estimate $\hat{I}$. The generator G is a U-net architecture with seven scales and six downsampling and upsampling stages. We compute the loss between the prediction $\hat{I}$ and the corresponding ground truth I_{ref} by

$$\mathcal{L}_c(I_{ref}, \hat{I}) = v_1\|\phi_l(\hat{I}) - \phi_l(I_{ref})\|_2 + v_2\|\hat{I} - I_{ref}\|_1, \tag{12}$$

where $v_1 = 0.5$ and $v_2 = 0.006$ are loss balancing weights and v_2 is added to keep the color fidelity, and ϕ_l extracts the feature maps from the l-th layer of pre-trained VGG-19. Specifically, we use the "conv3_3" layer of the VGG-19 network.

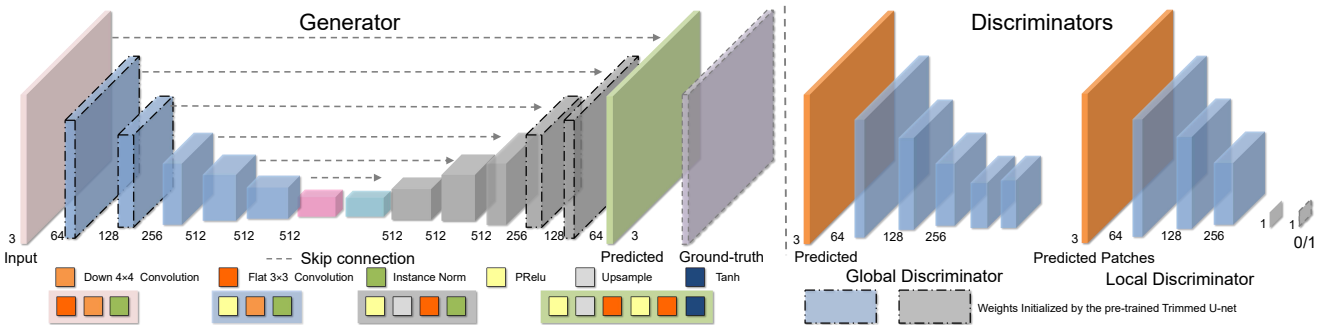

Fig. 3. Image reconstruction architecture. The generator model is a U-net architecture that has seven scales with six consecutive downsampling and upsampling operations. We adopt a global and a local discriminator to incorporate both full spatial contexts and local details. In addition, the layers marked by the dashed line is the trimmed U-net for the end-to-end designing stage, and they are initialized with the results of the designing stage.

Discriminators. As illustrated in Figure 3, we adopt a global discriminator to incorporate full spatial context and a local discriminator based on PatchGAN [Isola et al. 2017] to take advantage of local features. We adopt the relativistic "warping" on the least square GAN named RaGAN-LS loss [Kupyn et al. 2019] for a discriminator D can be expressed as:

$$\begin{aligned}\mathcal{L}_{adv}(x,z) = &\mathop{\mathbb{E}}_{x\sim\mathbb{P}_x}[(D(x) - \mathop{\mathbb{E}}_{z\sim\mathbb{P}_z}[(D(G(z)) - 1)^2] \\ &+ \mathop{\mathbb{E}}_{z\sim\mathbb{P}_z}[(D(G(z)) - \mathop{\mathbb{E}}_{\hat{x}\sim\mathbb{P}_{\hat{x}}}[(D(x) - 1)^2],\end{aligned} \tag{13}$$

where $\mathbb{P}_x$ and $\mathbb{P}_z$ are the distributions of the data and model, respectively. This proved faster and more stable than WGAN-GP [Arjovsky et al. 2017] in minimizing a model-generated image z and the ground truth x. The resulting total loss can be expressed as:

$$\mathcal{L}_{total} = \mathcal{L}_c(I_{ref}, \hat{I}) + \sigma_g L_{adv-g}(I_{ref}, \hat{I}) + \sigma_l L_{adv-l}(I_{ref}, \hat{I}), \tag{14}$$

where L_{adv-g} and L_{adv-l} represents global and local adversarial loss and $\sigma_g = \sigma_l = 0.01$.

4 IMPLEMENTATION AND PROTOTYPES

4.1 Datasets and Training details

For the training details of the end-to-end designing stage, please refer to Section 5 and Section 6 according to the requirements of the application LFOV and EDOF. With the lens parameters fixed, we train and finetune the image recovery network for both applications as follows. First, we rendered simulations with the full DIV2K dataset, and the texture plane are set according to the applications. We reserve the first 100 images in the DIV2K dataset [**?**] for quantitative comparisons, and use the remainder for training. Then we calibrate the lens distortion and find the homography to align the rendered result with the ground truth image. We use ADAM as the optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The learning rate is initialized to 10^{-4} for the first 50 epochs and linearly decayed to 0 over another 100 epochs using 256×256 patch pairs. All experiments were conducted using a single Nvidia RTX Titan GPU and each design takes around 20 hours.

4.2 Prototypes

Fabrication. Once the parameters c, κ and c_k of each lens profile are fixed after the end-to-end design, we fabricate each lens element with a coarse CNC machining process followed by a single-point diamond turning process. First, each lens blank was machined using a CNC machine with a precision of 0.05mm to prepare it for turning. Then we used a CNC machining system that supports 3-axis single point diamond turning (Nanotech 450) [Fang et al. 2013]. We use two substrates: polymethyl methacrylate (PMMA) with a refractive index of 1.493, and polycarbonate (PC) with a refractive index 1.5892, both measured at a principal wavelength of 550 nm. These materials represent a set of low index/low dispersion and high index/high dispersion materials that is required for designing achromatic optics.

Fig. 4. Prototypes and the rendered section views of our designed lenses. The top left shows our fabricated lens for LFOV (left) and EDOF (right) imaging, and the corresponding structures are shown in the medium/bottom left and medium/bottom right, respectively. The top right shows the assembled prototype with the camera body.

We consider two applications, large field-of-view (LFOV) and extended depth-of-field (EDOF). For the LFOV application, we design a lens system with two lens elements, made from PPMA and PC, respectively. For the EDOF application, we use three elements and six design surfaces, the corresponding materials are PMMA, PC, and PC.

System Integration. To demonstrate the proposed framework experimentally, we use a Sony A7 camera with 6,000×4,000 pixels and a pixel pitch of 5.96 μm. The equivalent focal length for both lens designs is 50mm, with aperture sizes of 12mm and 12mm for LFOV and EDOF, respectively. Correspondingly, both of the lens designs have f-numbers of about $F4$. The fabricated lenses are mounted by our custom-designed lens tubes, and both of them have a standard C-mount as shown in 4. Finally, both of the two lens tubes are mounted to the camera with a C/E mount adapter.

5 LARGE FIELD-OF-VIEW IMAGING

A modern complex system is effective in minimizing optical aberrations but the depth of the lens stack limiting in manufacturing high-quality LFOV lens with a low cost and will introduce additional issues, such as lens flare and complicated optical stabilization and assembling [Brady et al. 2012; Hasinoff and Kutulakos 2011; Venkataraman et al. 2013; Yuan et al. 2017]. In the last year, Peng *et al.* [2019] proposed the state-of-the-art of large FOV imaging with a thin-plate optics which adopts a virtual aperture design of two aspherical surfaces, reconstructed by a generative network. However, limited by the designing space of a single element, the PSFs at different FOVs are typically larger than 900 pixels yielding strong hazing and blurring artifacts recorded on the sensor. In addition, the optics and reconstruction network are not designed fully end-to-end, the recovered image left visible artifacts even with a powerful GAN recovery network.

With our proposed differentiable complex lens model, we can design a lens with multiple elements with an aspherical profile according to the needs of the applications and find the best compromise between the complexity of lens and image quality. To apply our framework to LFOV imaging ($\geq 30^\circ$), we set a texture plane at 1m away from the camera with the size around 437mm×54.5mm and set the sensor resolution to 4096×512 pixels to cover the full designed FOV. As the lens is designed rotationally sycemmatric, the full FOV should be calculated as $2\arctan(0.437/2*\sqrt{2}) = 34.3^\circ$. The pixel size in the simulation is defined as 6μm, matching the camera sensor used in the experiments. Limited by the GPU memory, we set the SPP to 64 for each rendering pass and average ten passes for a single scene to get a clean rendered image and corresponding accurate gradients. Refer to Section 3.1.2 for more details. To simulate a larger field of view with limited resources and reduce the time consumption, we align the sensor's left bottom corner with the optical center and simulate the image only in the first quadrant as the lens is symmetric.

We initialize the system with an initial lens design made of two lenses that are brought in focus on the optical axis. The materials are chosen as PMMA and PC for better cancelling of chromatic aberrations. To train the lens parameters for LFOV, we set all the conic coefficients κ of each surface as variables. As illustrated in Figure 2, a trimmed U-net architecture $\mathbf{G}_t$ connected to the end-to-end framework. The initial learning rate is set to 0.08 and 0.0008 for the lens parameters and network parameters, and both of them are decayed by a factor of 0.8 at each epoch. Our loss function is described as:

$$\mathcal{L}_c = \|\mathbf{G}_t(I_c) - I_{ref}\|_1. \tag{15}$$

Where the I_{ref} represents the ground truth. In addition, once the lens parameters are fixed, we take the parameters of the network to replace the corresponding layers (marked in Figure 3) of the image recovery network as the initialization and train the image recover network as described in Section 3 to process the images captured in the real experiments. In addition, the reference images are predistorted and resized at the beginning of each optimization step by the method described in Section 3.1.3 to make the image pairs matched pixel-to-pixel.

5.1 Evaluation in Simulation

Figure 5 shows a qualitative comparison of high-quality commercial available lenses and the state-of-the-art LFOV imaging lens (LFOV19) [Peng et al. 2019]. We also show the modulation transfer functions (MTFs) of each lens before and after the post-processing. Notice that some data in those MTF charts are missing due to the observed edges are heavily blurred and becomes uncalculatable. We first compare against the high-quality commercial available aspherical lens Thorlabs AL2550-A, which is optimized for focusing light incident on the aspherical side of the lens with minimal spherical aberration. Then we compared against an air-spaced doublet design ACA254-050-A, which provides superior spherical and chromatic aberration correction. As illustrated in Figure 5, the simulated PSFs by Zemax of AL2550-A and ACA254-050-A are well focused at the center FOV while corrupted when reaching a FOV 20°. The whole FOVs of simulated images shown in Figure 5 are all up to 30°. The Cooke triplet performs better compared with AL2550-A and ACA254-050-A but still has a noticeable blurry at a large angle. Refer to the supplementary material for more details. The state-of-the-art dual-surface aspherical lens design named LFOV19 has a better performance than the commercially available lenses as they balanced the aberrations of different FOV to achieve a larger FOV. However, this design yielding a very large PSF ($\geq$ 900 pixels) that overly degrades the image and has noticeable artifacts even with powerful generative post-processing. Our design, which is also compact and low-cost, introduces a differentiable ray-tracing based complex lens model that can directly optimize the lens parameters according to the tasks. The second column of Figure 5 illustrates that ours performs better from the PSF across the FOV. The third and fourth columns in Figure 5 give two examples of the cropped rendered simulation and corresponding reconstructed results, which use the same model and were retrained according to the lens. We show the cropped part of rendered simulation from a full FOV 0° (left side) to 30° (right side). Obviously, the results of AL2550-A and ACA254-050-A has a good performance at a small FOV but suffer from heavy blurring in off-axis regions. The LFOV19 shows an almost equal performance across the FOV but left noticeable artifacts. Ours has a better PSF behavior across the FOV, yielding better sensor measurements and reconstruction results.

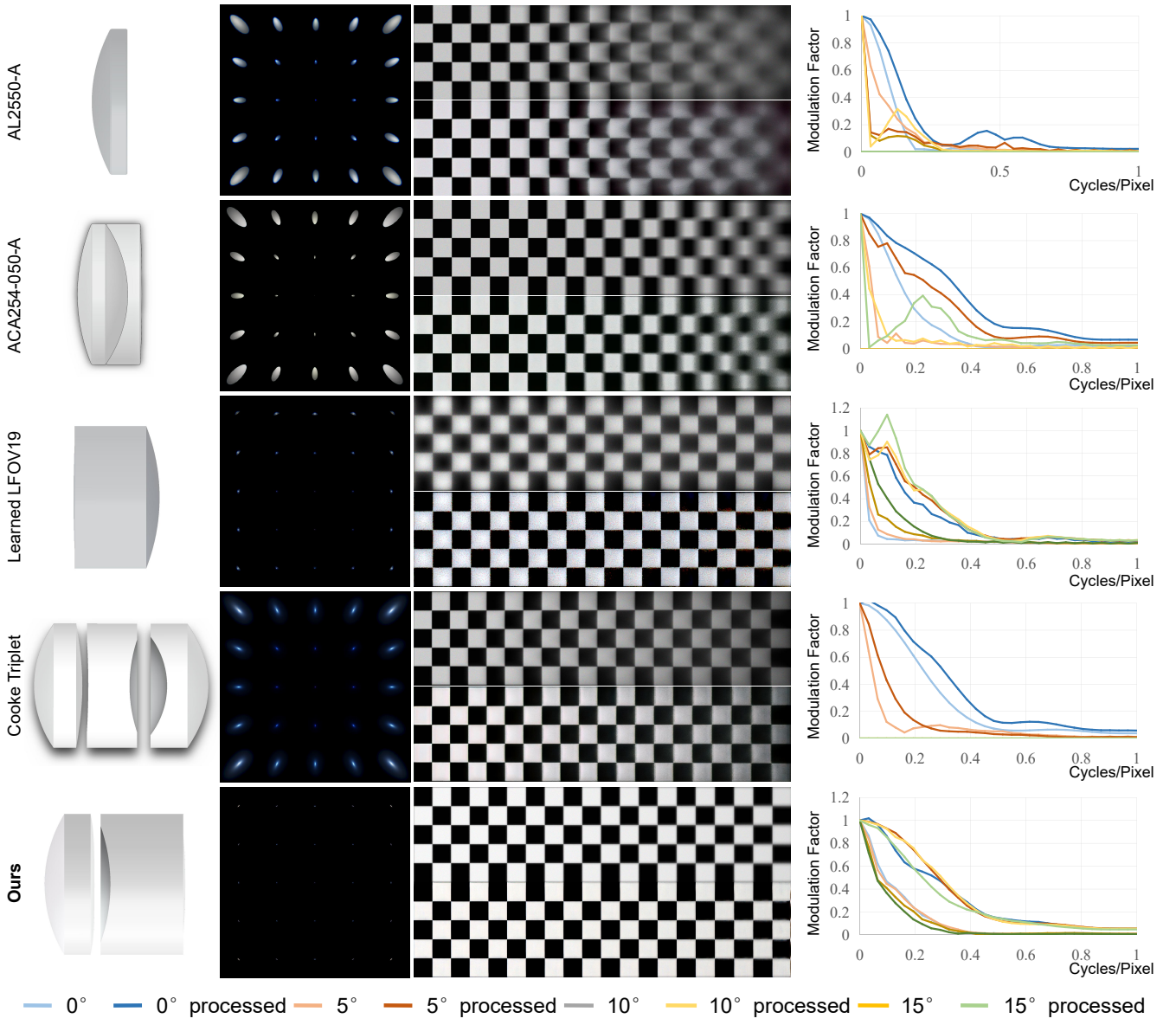

Fig. 5. Evaluation of LFOV imaging in simulation. We compare the performance of the state-of-the-art commercially available aspherical lens Thorlabs AL2550-A, lens pairs ACA254-050-A, jointly designed optics without modulation to flat lens [Peng et al. 2019], Cooke triplet and our end-to-end designed camera. All the texture planes are located at 1m away from the camera, and the simulations are based on ray optics without considering diffraction. The first column shows the section view of each lens, and the second column shows the corresponding PSFs at different angles up to 30˚. The third column shows the simulated sensor image (top) and recovered image (bottom). The fourth column shows the MTFs of each lens at different angles. The PSFs and rendered simulation of AL2550-A and ACA254-050-A lenses show a strong blurring at large angles. LFOV19 lens performs better in balancing PSF but left significant artifacts in both measurements and reconstructions. Cooke triplet performs better than AL2550-A and ACA254-050-A but still fails at a large FOV. Instead, our design shows a better PSF distribution, and the results have fewer artifacts. Notice that all lenses are adjusted to F4.

We also show the quantitative comparisons in simulation in Table 1. Obviously, our lens performs better in both PSNR and SSIM compared with the others over a FOV from 0° to 30°. Note that the training data and recovery network are re-rendered and retrained for each lens.

5.2 Experimental Results

To validate the practicability of the proposed differentiable complex lens model and the end-to-end framework, we fabricated the lens elements using single-point diamond turning and assembled them with the custom designed lens tube as shown in Figure 4. Figure 6 shows the pairs of “in-the-wild” captured raw sensor data (left) and

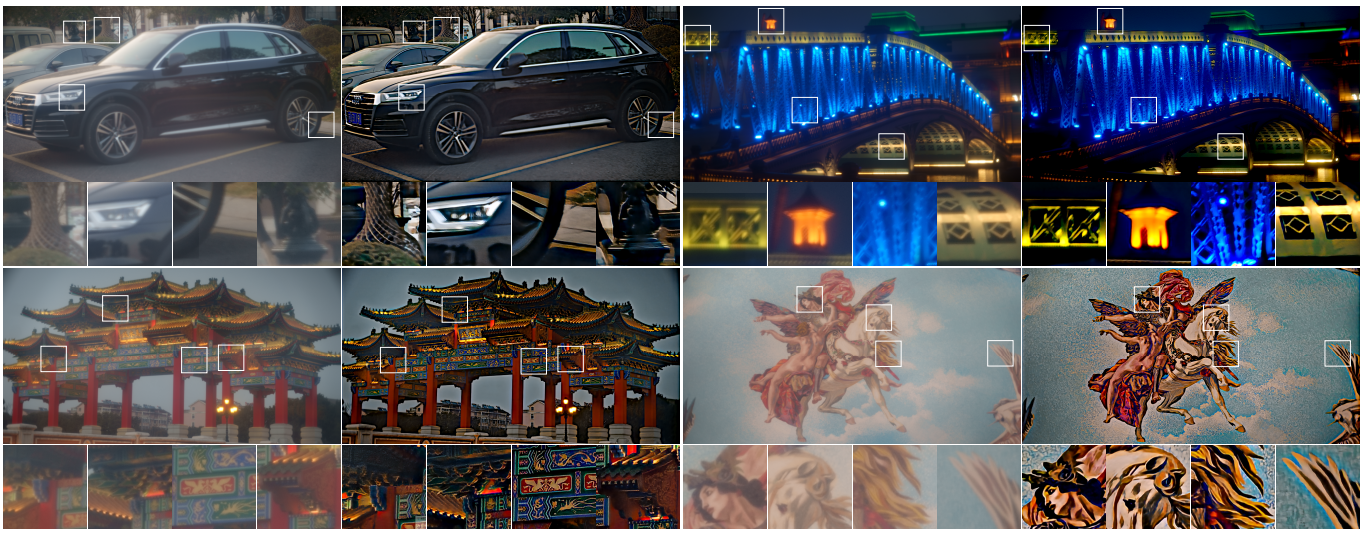

Fig. 6. Experimental results of large field of view imaging with two elements and four surfaces design. For each pair, we give the sensor measurement by our prototype camera and the reconstructed results. Please zoom in to see more details.

Table 1. Quantitative comparison of image recovery performance of different lenses. We compare PSNR values in dB and SSIM values over a FOV from 0° to 30°. Notice that all lenses are adjusted to F4.

	AL2550-A	ACA254-050A	LFOV19	Cooke	Ours
PSNR	16.96	19.03	16.86	15.724	**22.8**
SSIM	0.478	0.499	0.314	0.422	**0.719**

corresponding reconstructed results (right). The exposure times for each image are 33ms, 167ms, 167ms, 100ms with ISO 50. With our end-to-end designed imaging lens and reconstruction, we achieve a high-quality LFOV imaging with minor artifacts with only two lens elements. Notice the sensor measurements show haze artifacts, which is mainly introduced by the surface roughness, scratch, and low transparency of PMMA and PC in experiments. With our generative image reconstruction, we obtain clean results with fine details, as shown in Figure 6. Our lens design is compact and low-cost compared to commercial bulky lens and can get comparable results with the help of our differentiable complex lens model and end-to-end framework.

6 EXTENDED DEPTH OF FIELD

Computational EDOF cameras usually design an approximately depth-invariant PSF for one wavelength and then employ a simple deconvolution to the sensor capture to obtain an all-in-focus image [Cossairt and Nayar 2010; Cossairt et al. 2010; Dowski and Cathey 1995]. Recently, researchers proposed an end-to-end pipeline for diffractive optics or Zernike Basis [Sitzmann et al. 2018] and applied it to achromatic EDOF imaging. However, their optics model is based on the paraxial approximation, which is only a simple Fourier transform and can only deal with a single optical surface. With the proposed differentiable complex lens model and our end-to-end framework, we relax the designing space from a single surface to multiple surfaces for EDOF imaging.

To apply our end-to-end framework to EDOF imaging, we start with an initial triple-lens design with six surfaces where the second surface of the first and third elements are aspherical. This design is brought in good focus near the optical axis. The materials for the three lens elements are selected as PMMA, PC, and PC for better canceling of chromatic artifacts and easier fabrication. Refer to the supplementary material for more details. To obtain clean rendered images and corresponding accurate gradients despite the limited GPU memory, we use 10 rendering passes with 128 samples per pixel each. Please refer to Section 3.1.2 for more details.

We place the texture plane at 0.5m, 0.7m, 1m, and 1.5m away from the camera in simulation and try to find the best compromise between the different depths. The pixel size in the simulation is set to 6μm and the sensor resolution is set 256×256 pixels while the texture plane sizes are set to 13.82mm×13.82mm, 20.51mm×20.51mm, 30.72mm×30.72mm and 46.08mm×46.08mm, respectively. To train the lens parameters to achieve a EDOF camera, we set the conic coefficients κ of the spherical surface as the variable (four surfaces in total). The initial learning rate is set to 0.08 for the lens parameters, and they are decayed by a factor of 0.8 at each epoch. Our loss function can be expressed as:

$$\mathcal{L}_c = \Sigma_j \zeta_i \|I_{ci} - I_{c2}\|_1 + \Sigma_j 3(1 - SSIM(I_c, I_{ref})), \qquad (16)$$

where ζ_i represents the weights for different depths, and we set them to $\{3, 3, 0.3, 1\}$ corresponding to the depths mentioned above, respectively. I_{c2} represents the sensor measurement with the texture plane placed a distance of 1m, and we take it as a reference to balancing the blurring amount over different depths. For each depth, we adopt a structural similarity index measure (SSIM) loss between the sensor measurement and the corresponding clean reference as the brightness and gamma might mismatch with the reference. In addition, the reference images are pre-distorted and resized at the beginning of each optimization step by the method described in Section 3.1.2 to align the image pairs pixel-to-pixel.

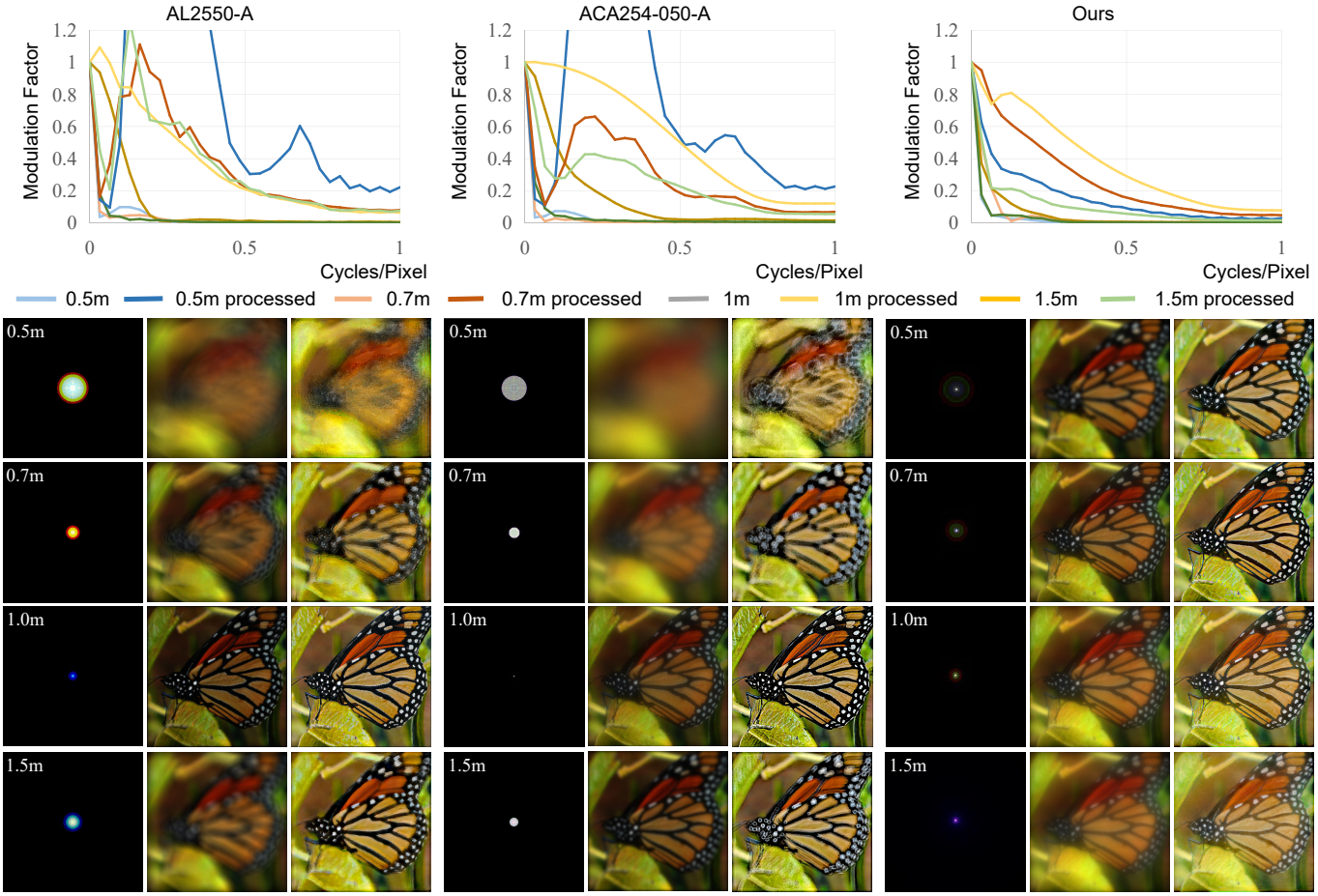

Fig. 7. Evaluation of extended depth of field imaging in simulation. We compare the performance of the state-of-the-art commercially available aspherical lenses, including Thorlabs AL2550-A and ACA254-050-A. The first row shows the MTFs of each lens before and after post-processing at different depths. All the texture planes are placed 1m away from the camera, and the simulation is based on ray optics without considering diffraction. The second row shows the corresponding PSFs at the selected depth. The third column shows the simulated sensor image. Obviously, the PSFs of rendered simulation of AL2550-A and ACA254-050-A lenses exhibit a strong blur when out of focus. Instead, our design shows an almost depth invariant PSF and results with fewer artifacts. Additional results are available in the supplementary material. Notice that all lenses are adjusted to F4.

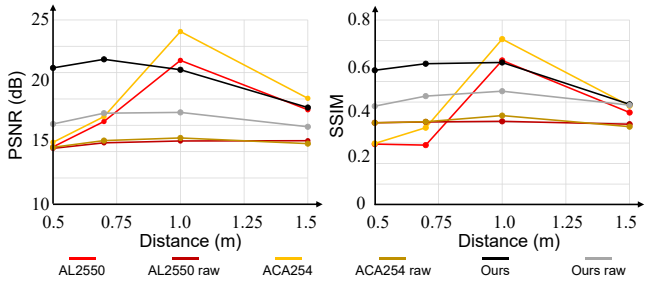

Fig. 8. Quantitative comparison of image recovery performance of different lenses. We compare PSNR values in dB and SSIM values at 0.5m, 0.7m, 1.0m, and 1.5m. Notice that all lenses are adjusted to F4.

6.1 Evaluation in Simulation

We first validate our lens design in simulation and compared our lens with the high-quality commercially available lenses, including AL2550-A and ACA254-050-A. We focus all the lenses at 1m away from the camera. As illustrated in Figure 7, the simulated center PSFs by Zemax of AL2550-A and ACA254-050-A behave well when in focus. However, their PSF becomes unacceptably large when out of focus. In contrast, our design has an almost depth invariant PSF behavior compared with the others. We first validate our lens design in simulation and compared our lens with the high-quality commercially available lenses, including AL2550-A and ACA254-050-A. We focus all the lenses at 1m away from the camera. As illustrated in Figure 7, the simulated center PSFs by Zemax of AL2550-A and ACA254-050-A behave well when in focus. However, their PSF becomes unacceptably large when out of focus. In contrast, our design has an almost depth invariant PSF behavior compared with the

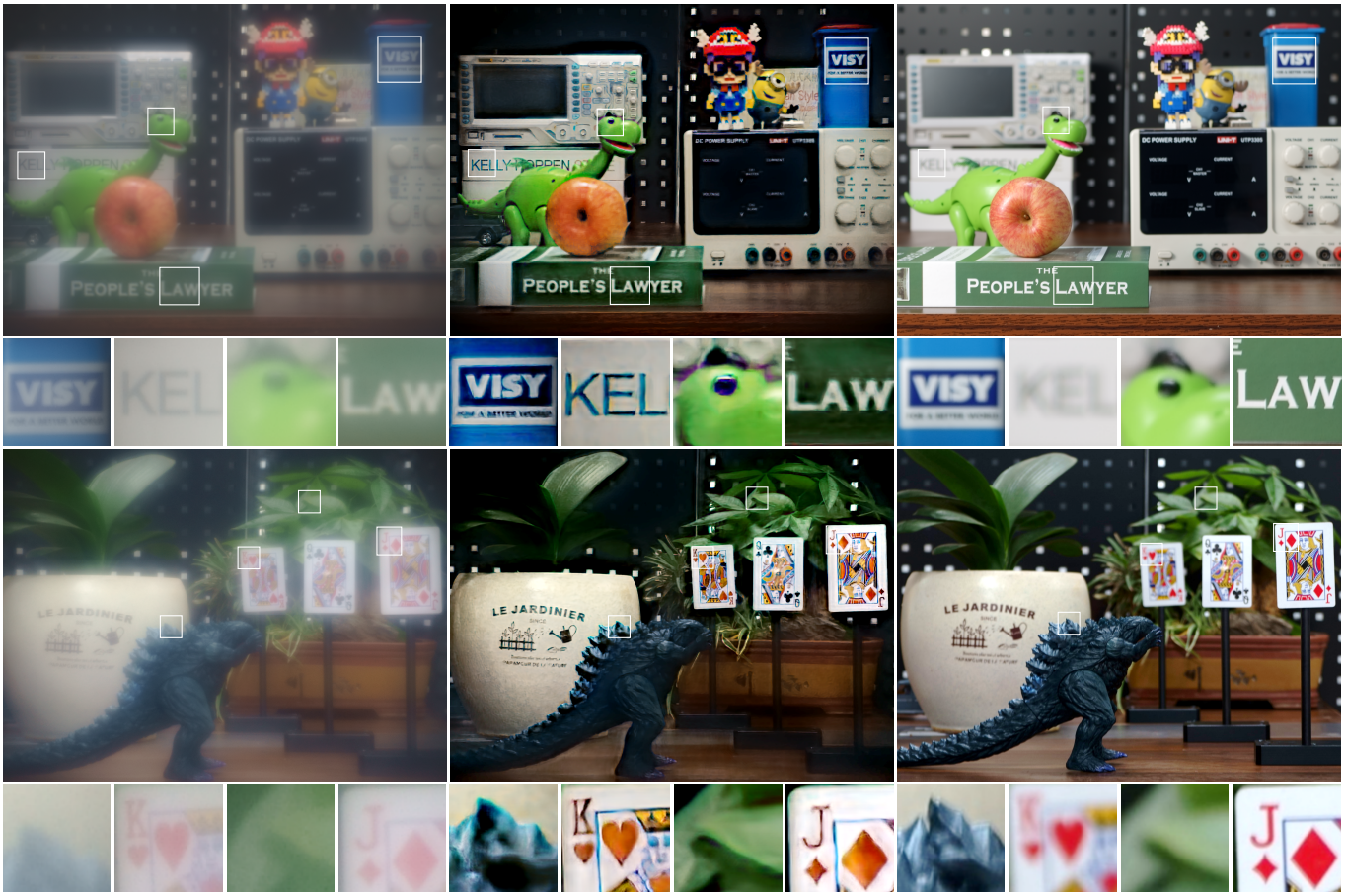

Fig. 9. Experimental results of extended depth of field with three elements and six surfaces design. The left column shows the raw sensor data from our design, the center column shows our reconstruction result and the right column shows images captured by a commercial Sony 28-70mm zoom lens adjusted to 50mm/F4.5. The objects shown in these two figures are placed from around 0.8m to 1.8m, and we succeed in obtaining the all-in-focus image. Please zoom in to see more details.

others. Besides, the MTFs in Figure 7 show that the MTF of our optimized lens is closer to the desired MTF in optical systems: smoothly and monotonously decreasing from an amplitude of 100% for the DC term to ca. 10% at the Nyquist limit of the SR image, with no erroneous maxima for higher frequencies. Instead, the others show an obvious outlier for the post-processed data and worse performance before processing.

We further rendered the scene at different depths for each lens to provide further evidence that our end-to-end design has a larger DOF. We rendered the whole dataset for each rendered scene and retrained the recovery network for fair comparison for each rendered scene and recovered estimation pairs. As illustrated in Figure 7, AL2550-A and ACA254-050-A have better performance when in focus for both rendered results and corresponding recoveries but break when out of focus. In contrast, our design has a depth-balanced performance in both sensor measurements and reconstructed images.

We also show the quantitative comparisons in simulation in Figure 8. Our lens performs better balancing over depth in both PSNR and SSIM compared with the others. For a fair comparison, all lenses are adjusted with an aperture of F4. Note that the training data and recovery network are re-rendered and retrained according to each lens. Furthermore, the rendered images' energy distribution might vary with the lenses and make them different from the reference images, causing a relatively low matric value and less accuracy.

6.2 Experimental Results

To demonstrate the practicability of our approach in EDOF, we fabricated and assembled the lenses with the custom-designed lens tube as shown in Figure 4. Figure 9 shows the captured raw sensor measurement (left), reconstructed results (middle), and the reference image captured by a Sony 28-70mm standard zoom lens adjusted to 50mm/F4.5. The exposure times for each image are all set to 200ms with ISO 50. Figure 9 illustrates that we achieved good performance and high image quality over a large DOF. Compared with our lens (F4), the Sony 28-70mm standard zoom lens has a larger f-number but worse DOF performance. Notice the sensor measurements show haze artifacts, which has been discussed in Section 5.2 and Section 7.

7 DISCUSSION AND CONCLUSION

7.1 Discussion

Stablity and efficiency. We have introduced a differentiable complex lens model that can be connected with tailored image reconstructions. Compared to conventional lens design, which requires much experience in setting up merit functions to affect the desired design characteristics, our approach reduces the need for human involvement. However, both methods can converge to a local minimum if the starting point of the design is too far off from a feasible solution. Like traditional lens design, which requires a proper initialization, our approach still requires a good initial structure that can then be further optimized automatically. Our data-driven optics do not yet take into account many standard tasks of optical design, such as zoom and focus changes, or design aspects such as tolerancing or anti-reflective coatings. We believe, however, that such extensions will be easy to add to the framework in the future.

Our approach traces hundreds of rays for each pixel as for the computational efficiency, resulting in millions of rays to compute the gradients. This is less time-efficient even with the help of the state-of-the-art ray tracing cores and has a vast space to optimize. In future work, we would like to introduce a patch wised rendering strategy instead of tracing the whole FOV to improve the computational efficiency during the designing stage.

Fabrication. To demonstrate our differentiable optics model and end-to-end the pipeline, we fabricated two prototypes for different applications using single-point diamond turning from PC and PMMA material. PC and PMMA are easy to manufacture, but the stability, transparency, and the easily transformed make it hard to achieve a good imaging quality for a complex lens system. Besides, the center alignment of the lenses and surfaces is a challenging task during machining and assembling. As a result, the real captured sensor images have haze artifacts compared with the simulations. However, many of these issues can likely be overcome in mass production, such as injection molding fabrication processes like the ones already used in the manufacture of cell phone cameras. We also would like to fabricate lenses from optical glass with coatings to reduce the stray light in the future.

7.2 Conclusion

We proposed a novel differentiable complex model that provides a new approach for optics designing and an end-to-end framework that can be tailored for specifics tasks. We demonstrated our model and pipeline on two applications, including LFOV and EDOF imaging with compact lens designs, and tested both in real-world experiments. In addition, our model can not only be applied to the end-to-end designing of optics but also offers a new approach for simulating lens' aberrations, which makes it less cumbersome to obtain a large, well-aligned training dataset for the image recovery training stage. While the proposed approach enables practical, high-quality imagery with compact designs, stability to initialization, and computational efficiency need to be further investigated in future work.

In the future, it might also be interesting to explore hybrid refractive/diffractive optical systems, and to incorporate features like coatings and other optical effects. Furthermore, building a knowledge graph that contains a large library of classic designs are an exciting direction to get rid of human involvement in initializing our system and making the design process fully automatic.

ACKNOWLEDGEMENTS

This work was supported by KAUST baseline funding. We would like to thank Merlin Nimier-David and Wenzel Jakob for discussions on an earlier version of this idea and for help with the Mitsuba 2 system, and Guang Li from Point Spread Technology for mechanical support. Finally we would like to thank the Siggraph reviewers for their valuable comments.

REFERENCES

Nick Antipa, Grace Kuo, Reinhard Heckel, Ben Mildenhall, Emrah Bostan, Ren Ng, and Laura Waller. 2018. DiffuserCam: lensless single-exposure 3D imaging. *Optica* 5, 1 (2018), 1–9.

Martin Arjovsky, Soumith Chintala, and Léon Bottou. 2017. Wasserstein Generative Adversarial Networks. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70* (Sydney, NSW, Australia) *(ICML'17)*. JMLR.org, 214–223.

Seung-Hwan Baek, Hayato Ikoma, Daniel S Jeon, Yuqi Li, Wolfgang Heidrich, Gordon Wetzstein, and Min H Kim. 2020. End-to-end hyperspectral-depth imaging with learned diffractive optics. *arXiv preprint arXiv:2009.00463* (2020).

Sai Bangaru, Tzu-Mao Li, and Frédo Durand. 2020. Unbiased Warped-Area Sampling for Differentiable Rendering. *ACM Transactions on Graphics (TOG)* 39, 6 (2020), 245:1–245:18.

V. Boominathan, J. K. Adams, J. T. Robinson, and A. Veeraraghavan. 2020. PhlatCam: Designed Phase-Mask Based Thin Lensless Camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 42, 7 (2020), 1618–1629.

David J Brady, Michael E Gehm, Ronald A Stack, Daniel L Marks, David S Kittle, Dathon R Golish, EM Vera, and Steven D Feller. 2012. Multiscale gigapixel photography. *Nature* 486, 7403 (2012), 386.

W Thomas Cathey and Edward R Dowski. 2002. New paradigm for imaging systems. *Applied Optics* 41, 29 (2002), 6080–6092.

Ayan Chakrabarti. 2016. Learning Sensor Multiplexing Design through Back-propagation. In *Advances in Neural Information Processing Systems*, D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett (Eds.), Vol. 29. Curran Associates, Inc.

Julie Chang and Gordon Wetzstein. 2019a. Deep Optics for Monocular Depth Estimation and 3D Object Detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*.

Julie Chang and Gordon Wetzstein. 2019b. Deep optics for monocular depth estimation and 3d object detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 10193–10202.

Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. 2018. Learning to See in the Dark. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2018).

Shane Colburn, Alan Zhan, and Arka Majumdar. 2018. Metasurface optics for full-color computational imaging. *Science Advances* 4, 2 (2018).

Oliver Cossairt and Shree Nayar. 2010. Spectral focal sweep: Extended depth of field from chromatic aberrations. In *IEEE International Conference on Computational Photography (ICCP)*. IEEE, 1–8.

O. Cossairt, C. Zhou, and S.K. Nayar. 2010. Diffusion Coding Photography for Extended Depth of Field. *ACM Transactions on Graphics (TOG)* (Aug 2010).

O. S. Cossairt, D. Miau, and S. K. Nayar. 2011. Gigapixel Computational Imaging. In *IEEE International Conference on Computational Photography (ICCP)*. 1–8.

Geoffroi Côté, Jean-François Lalonde, and Simon Thibault. 2019. Extrapolating from lens design databases using deep learning. *Opt. Express* 27, 20 (Sep 2019), 28279–28292.

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.

Paul E. Debevec and Jitendra Malik. 1997. Recovering High Dynamic Range Radiance Maps from Photographs. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '97)*. ACM Press/Addison-Wesley Publishing Co., USA, 369–378.

Edward R Dowski and W Thomas Cathey. 1995. Extended depth of field through wave-front coding. *Applied optics* 34, 11 (1995), 1859–1866.

Xiong Dun, Hayato Ikoma, Gordon Wetzstein, Zhanshan Wang, Xinbin Cheng, and Yifan Peng. 2020. Learned rotationally symmetric diffractive achromat for full-spectrum computational imaging. *Optica* 7, 8 (Aug 2020), 913–922.

FZ Fang, XD Zhang, A Weckenmann, GX Zhang, and C Evans. 2013. Manufacturing and measurement of freeform optics. *CIRP Annals* 62, 2 (2013), 823–846.

Angel Flores, Michael R. Wang, and Jame J. Yang. 2004. Achromatic hybrid refractive-diffractive lens with extended depth of focus. *Applied Optics* 43, 30 (Oct 2004), 5618–5630.

Grant R Fowles. 2012. *Introduction to modern optics.* Courier Dover Publications.

Qi Guo, Iuri Frosio, Orazio Gallo, Todd Zickler, and Jan Kautz. 2018. Tackling 3D ToF Artifacts Through Learning and the FLAT Dataset. In *The European Conference on Computer Vision (ECCV)*. Springer.

Harel Haim, Shay Elmalem, Raja Giryes, Alex Bronstein, and Emanuel Marom. 2018. Depth Estimation From a Single Image Using Deep Learned Phase Coded Mask. *IEEE Transactions on Computational Imaging* 4 (2018), 298–310.

Samuel W Hasinoff and Kiriakos N Kutulakos. 2011. Light-efficient photography. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33, 11 (2011), 2203–2214.

Felix Heide, Qiang Fu, Yifan Peng, and Wolfgang Heidrich. 2016. Encoded diffractive optics for full-spectrum computational imaging. *Scientific Reports* 6 (2016).

Roarke Horstmeyer, Richard Y. Chen, Barbara Kappes, and Benjamin Judkewitz. 2017. Convolutional neural networks that teach microscopes how to image. *ArXiv* abs/1709.07223 (2017).

Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. 2017. Image-to-Image Translation with Conditional Adversarial Networks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2017).

Francis A Jenkins and Harvey E White. 2018. *Fundamentals of optics.* Tata McGraw-Hill Education.

Daniel S. Jeon, Seung-Hwan Baek, Shinyoung Yi, Qiang Fu, Xiong Dun, Wolfgang Heidrich, and Min H. Kim. 2019. Compact Snapshot Hyperspectral Imaging with Diffracted Rotation. *ACM Transactions on Graphics (TOG)* 38, 4 (2019), 117:1–13.

Michael Kellman, Emrah Bostan, Michael Chen, and Laura Waller. 2019. Data-Driven Design for Fourier Ptychographic Microscopy. In *IEEE International Conference on Computational Photography (ICCP)*. IEEE, 1–8.

Salman S. Khan, Adarsh V. R. , Vivek Boominathan, Jasper Tan, Ashok Veeraraghavan, and Kaushik Mitra. 2019. Towards Photorealistic Reconstruction of Highly Multiplexed Lensless Images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*.

Craig Kolb, Don Mitchell, and Pat Hanrahan. 1995. A realistic camera model for computer graphics. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*. 317–324.

Alankar Kotwal, Anat Levin, and Ioannis Gkioulekas. 2020. Interferometric Transmission Probing with Coded Mutual Intensity. 39, 4, Article 74 (July 2020), 16 pages.

Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiri Matas. 2017. DeblurGAN: Blind Motion Deblurring Using Conditional Adversarial Networks. *arXiv preprint arXiv:1711.07064* (2017).

Orest Kupyn, Tetiana Martyniuk, Junru Wu, and Zhangyang Wang. 2019. DeblurGAN-v2: Deblurring (Orders-of-Magnitude) Faster and Better. In *The IEEE International Conference on Computer Vision (ICCV)*.

Anat Levin. 2010. Analyzing Depth from Coded Aperture Sets. In *Computer Vision – ECCV 2010*, Kostas Daniilidis, Petros Maragos, and Nikos Paragios (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 214–227.

Anat Levin, Rob Fergus, Frédo Durand, and William T. Freeman. 2007. Image and Depth from a Conventional Camera with a Coded Aperture. *ACM Transactions on Graphics (TOG)* 26, 3 (July 2007), 70–es.

Anat Levin, Samuel W Hasinoff, Paul Green, Frédo Durand, and William T Freeman. 2009. 4D frequency analysis of computational cameras for depth of field extension. In *ACM Transactions on Graphics (TOG)*, Vol. 28. ACM, 97.

Zhiqiang Liu, Angel Flores, Michael R. Wang, and Jianwen J. Yang. 2007. Diffractive infrared lens with extended depth of focus. *Optical Engineering* 46, 1 (2007), 1 – 9.

Daniel Malacara-Hernández and Zacarías Malacara-Hernández. 2016. *Handbook of optical design.* CRC Press.

S. Mann and Rosalind W. Picard. 1994. Being 'undigital' with digital cameras: extending dynamic range by combining differently exposed pictures.

Christopher A Metzler, Hayato Ikoma, Yifan Peng, and Gordon Wetzstein. 2020. Deep optics for single-shot high-dynamic-range imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1375–1385.

Mehjabin Monjur, Leonidas Spinoulas, Patrick R Gill, and David G Stork. 2015. Ultra-miniature, computationally efficient diffractive visual-bar-position sensor. In *Proc. SensorComm*. IEIFSA.

Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. 2017. Deep multi-scale convolutional neural network for dynamic scene deblurring. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 1, 2, 3.

S.K. Nayar, V. Branzoi, and T. Boult. 2004. Programmable Imaging using a Digital Micromirror Array. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* I, 436–443.

Elias Nehme, Daniel Freedman, Racheli Gordon, Boris Ferdman, Tomer Michaeli, and Yoav Shechtman. 2019. Dense three dimensional localization microscopy by deep learning.

Merlin Nimier-David, Delio Vicini, Tizian Zeltner, and Wenzel Jakob. 2019. Mitsuba 2: A Retargetable Forward and Inverse Renderer. *ACM Transactions on Graphics (TOG)* 38, 6 (Dec. 2019).

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 Transactions on Graphics (TOG)* 38, 6, Article 219 (Nov. 2019), 14 pages.

E. Reinhard and K. Devlin. 2005. Dynamic range reduction inspired by photoreceptor physiology. *IEEE Transactions on Visualization and Computer Graphics* 11, 1 (2005), 13–24.

M. Rouf, R. Mantiuk, W. Heidrich, M. Trentacoste, and C. Lau. 2011. Glare Encoding of High Dynamic Range Images. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Christian J. Schuler, Harold Christopher Burger, Stefan Harmeling, and Bernhard Scholkopf. 2013. A Machine Learning Approach for Non-blind Image Deconvolution. In *Proc. Computer Vision and Pattern Recognition*.

Yoav Shechtman, Lucien E Weiss, Adam S. Backer, Maurice Y. Lee, and W E Moerner. 2016. Multicolour localization microscopy by point-spread-function engineering. *Nature photonics* 10 (2016), 590–594.

Yichang Shih, Brian Guenter, and Neel Joshi. 2012. Image enhancement using calibrated lens simulations. In *European Conference on Computer Vision (ECCV)*. Springer, 42–56.

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 (TOG)* 37, 4 (2018), 114.

Warren J. Smith. 2005. *Modern lens design.* McGraw-Hill.

David G Stork and Patrick R Gill. 2013. Lensless ultra-miniature CMOS computational imagers and sensors. *Proc. SENSORCOMM* (2013), 186–190.

David G Stork and Patrick R Gill. 2014. Optical, mathematical, and computational foundations of lensless ultra-miniature diffractive imagers and sensors. *International Journal on Advances in Systems and Measurements* 7, 3 (2014), 4.

Qilin Sun, Xiong Dun, Yifan Peng, and Wolfgang Heidrich. 2018. Depth and Transient Imaging With Compressive SPAD Array Cameras. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Qilin Sun, Ethan Tseng, Qiang Fu, Wolfgang Heidrich, and Felix Heide. 2020a. Learning Rank-1 Diffractive Optics for Single-Shot High Dynamic Range Imaging. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Qilin Sun, Jian Zhang, Xiong Dun, Bernard Ghanem, Yifan Peng, and Wolfgang Heidrich. 2020b. End-to-End Learned, Optically Coded Super-Resolution SPAD Camera. *ACM Transactions on Graphics (TOG)* 39, 2, Article 9 (March 2020), 14 pages.

Sara C Tucker, W Thomas Cathey, and Edward R Dowski. 1999. Extended depth of field and aberration control for inexpensive digital microscope systems. *Optics Express* 4, 11 (1999), 467–474.

Kartik Venkataraman, Dan Lelescu, Jacques Duparré, Andrew McMahon, Gabriel Molina, Priyam Chatterjee, Robert Mullis, and Shree Nayar. 2013. Picam: An ultra-thin high performance monolithic camera array. *ACM Transactions on Graphics (TOG)* 32, 6 (2013), 166.

Yicheng Wu, Vivek Boominathan, Huaijin Chen, Aswin Sankaranarayanan, and Ashok Veeraraghavan. 2019a. PhaseCam3D – Learning Phase Masks for Passive Single View Depth Estimation. In *IEEE International Conference on Computational Photography (ICCP)*.

Y. Wu, V. Boominathan, H. Chen, A. Sankaranarayanan, and A. Veeraraghavan. 2019b. PhaseCam3D â€" Learning Phase Masks for Passive Single View Depth Estimation. In *IEEE International Conference on Computational Photography (ICCP)*. IEEE Computer Society, Los Alamitos, CA, USA, 1–12.

Y. Wu, F. Li, F. Willomitzer, A. Veeraraghavan, and O. Cossairt. 2020. WISHED: Wavefront imaging sensor with high resolution and depth ranging. In *IEEE International Conference on Computational Photography (ICCP)*. 1–10.

Li Xu, Jimmy SJ Ren, Ce Liu, and Jiaya Jia. 2014. Deep convolutional neural network for image deconvolution. In *Advances in Neural Information Processing Systems*. 1790–1798.

Xiaoyun Yuan, Lu Fang, Qionghai Dai, David J Brady, and Yebin Liu. 2017. Multiscale gigapixel video: A cross resolution image matching and warping approach. In *IEEE International Conference on Computational Photography (ICCP)*. IEEE, 1–9.

Cheng Zhang, Bailey Miller, Kai Yan, Ioannis Gkioulekas, and Shuang Zhao. 2020. Path-Space Differentiable Rendering. *ACM Transactions on Graphics (TOG)* 39, 4 (2020), 143:1–143:19.

Cheng Zhang, Lifan Wu, Changxi Zheng, Ioannis Gkioulekas, Ravi Ramamoorthi, and Shuang Zhao. 2019. A Differential Theory of Radiative Transfer. *ACM Transactions on Graphics (TOG)* 38, 6 (2019), 227:1–227:16.

Jiawei Zhang, Jinshan Pan, Wei-Sheng Lai, Rynson WH Lau, and Ming-Hsuan Yang. 2017. Learning fully convolutional networks for iterative non-blind deconvolution. (2017).

Xuaner Zhang, Ren Ng, and Qifeng Chen. 2018. Single Image Reflection Separation with Perceptual Losses. In *IEEE Conference on Computer Vision and Pattern Recognition*.