ELEC 549: End-to-end Optimization of Optics and Image Processing

Krish Kabra[a1] **and Yuhao Liu**[a2]

[a]Graduate Fellow, Rice University

December 17, 2021

Traditional optical design has long been driven by measurement rather than application. Computational cameras have emerged as a means to design domain-specific cameras with a co-design of the system's optics and image processing algorithms, yet many approaches are designed based on heuristics or proxy metrics on the point spread function. Ideally, the optimal computational camera for a given task requires the joint optimization of parameters from both the image forming optics and sensor-data processing algorithm in an end-to-end fashion. Following the success of machine learning-based image processing algorithms, Sitzmann *et al.* propose a framework in which the design of a single lens imaging system is learnt in an end-to-end fashion using a high-resolution natural scene image dataset and a differential image formation model. In this report, we utilize the proposed end-to-end optics and image processing framework to learn the design for achromatic extended depth of field imaging system.

computational imaging | end-to-end optimization

1. Introduction

Modern imaging technologies have given us powerful capabilities to capture the world on devices ranging greatly in form factor, from handheld to millimeter-length. Moreover, with recent advances in image processing algorithms, modern vision systems can accomplish a variety of tasks in challenging environments, such as low-lighting (1) and poor weather (2). However, while image processing algorithms have been optimized for specific tasks, commercial optical design has largely focused on improving measurement through the reduction of aberrations i.e. deviations from Gauss's linear model of optics (3). This general-purpose design approach has been successful, yet leaves an important unanswered question: *What is the optimal camera design for a given task?*

Computational cameras (4), domain-specific camera systems designed with the joint consideration of both optics and computational processing techniques to reconstruct a final image, have emerged over the last two decades to address this question. Raskar *et al.* rapidly flutter the camera's shutter open and closed during an exposure with a binary pseudo-random sequence, thereby preserving high frequency details that would have otherwise been lost in a single long exposure and allowing the corresponding deconvolution to become a well-posed problem (5). Cossairt *et al.* introduce a custom designed diffuser into their camera system as to engineer a depth-invariant point spread function (PSF) to enable extended depth-of-field (EDoF) imaging. However, these approaches are either heuristic or use some proxy metric on the PSF rather than considering the image quality after post-processing. Ideally, the optimal computational camera for a given task requires the joint optimization of parameters from both the image forming optics and sensor-data processing algorithm in some end-to-end fashion.

Sitzmann *et al.* (6) proposed a framework in which this exact end-to-end optimization of optics and image processing could be achieved for a singular refractive or diffractive optical element. Following the success of machine learning-based image processing algorithms that utilize large image datasets for parameter optimization, the end-to-end framework jointly optimizes optical parameters and image reconstruction parameters using such datasets by minimizing the deviation between the true and reconstructed images. The optimization is made possible through the development of a fully differentiable wave optics image formation pipeline and stochastic gradient methods. An overview of the framework can be seen visually in Figure 1.

In this report, we outline the methods used by Sitzmann *et al.* and attempt to reproduce their results. Our objectives are:

- To re-implement the proposed end-to-end optimization framework for a singular optical element system. *
- To validate the framework in simulation for the applications of achromatic imaging and achromatic EDoF.

2. End-to-end Optimization of Optics and Reconstruction

A. Image Formation Model. To build the image formation model, a wave-based model that accounts for diffraction and wavelength-dependent effects, derived from Fourier optics (7), is utilized.

A.1. Wave-based point spread function. A single refractive or diffractive optical element, such as a thin lens, imparts phase delays to an incoming wave field proportional to its thickness $h(x, y)$:

$$\phi_d(x, y) = \frac{2\pi\Delta n}{\lambda} h(x, y) \quad [1]$$

Here, λ is the wavelength in the input field and Δn is the difference in refractive index between air and the material of the optical element.

A wave field $U_\lambda(x, y, z = 0)$ with amplitude $A(x, y)$ and phase $\phi(x, y)$ incident on the optical element will then be affected as follows:

$$U_\lambda(x, y, z = 0) = A(x, y)e^{j(\phi_d(x,y)+\phi(x,y))} \quad [2]$$

After passing through the optical element, the wave propagates to the image sensor location a distance z away. This propagation can be approximated by a Fresnel propagation

* Re-implemented source code is available at https://github.com/Way-Yuhao/End2End-Optics

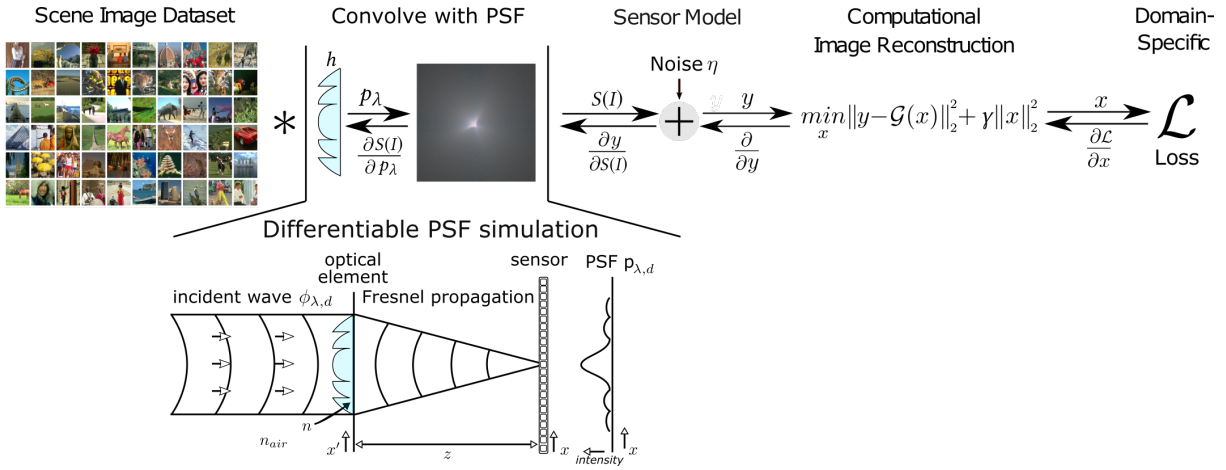


Fig. 1. An end-to-end optics and image processing framework. A large image dataset of natural scenes is passed through a fully differentiable wave-optics model by convolving a batch of images with a simulated PSF p of the current optical system. A sensor model is simulated by adding sensor read noise eta. Computational image reconstruction is achieved by solving a Tikhonov-regularized least-squares problem with the optical image formation model $\mathcal{G}$. Finally a differentiable loss $\mathcal{L}$, such as men-squared error with respect to the ground-truth image, is defined on the reconstructed images. The error can then be backpropagated all the way back to the optical element. Credits: Image from Sitzmann *et al.* (6)

operator, which is reasonable given that we are within the paraxial regime i.e. $z >> \lambda$. Thus, the wave field at the sensor plane is:

$$U_\lambda(x,y,z) = \frac{e^{jkz}}{j\lambda z}\iint U(x,y,0)e^{\frac{jk}{2z}\left((x-x')^2+(y-y')^2\right)}dx'dy' \quad [3]$$

Here, $k = 2\pi/\lambda$ is the wavenumber.

Finally, to derive the point spread function (PSF) $p(x,y)$ of the overall system, we can simulate an input wave field that corresponds to a single point source. For a simple PSF with no depth dependence, we can place the point source at optical infinity, which corresponds to an input wave field that is a plane wave i.e. setting phase $\phi(x,y) = 0$. this leads to the following result:

$$p_\lambda(x,y) \propto \left|\mathcal{F}\left\{A(x,y)e^{j\phi_d(x,y)}e^{\frac{j\pi}{\lambda z}(x^2+y^2)}\right\}\right|^2 \quad [4]$$

We can approximate a depth dependent PSF by simulating an spherical input wave field at a distance d with the appropriate curvature i.e. setting phase $\phi(x,y) = \sqrt{(x^2+y^2)+d^2}$. In this case, the PSF is no longer a shift-invariant convolution. Nevertheless, for sufficiently large distances between the scene and the optical element, a shift-invariant formation model may still be a good approximation (8), although off-axis aberrations like coma will be neglected.

A.2. From PSF to image. Assuming the PSF is shift-invariant operation, the image at the sensor plane can be obtained by means of a convolution. Accounting for the wavelength sensitivity κ_c of the sensor for each of the color channels R, G, B, we find:

$$I_c(x,y) = \int (I_\lambda * p_\lambda)(x,y)\kappa_c(\lambda)d \quad [5]$$

Here, I_λ is the original wavelength dependent scene, and c denotes the color channel.

A.3. Sensor. The image $I_c(x,y)$ formed on the sensor is integrated over the sensor pixels and corrupted by noise, yielding a measurement y_c:

$$y_c = S(I_c) + \eta \quad [6]$$

where S is the pixel integration and sampling operator and $\eta \sim \mathcal{N}(0,\sigma^2)$ is Gaussian read noise.

A.4. Image reconstruction. The final stage the proposed model is image reconstruction. We reconstruct the estimate $\tilde{I}_c$ of the source image I_λ by solving the Tikhonov regularized least-squares problem:

$$\min_{\{\tilde{I}\}} ||y_c - S(\tilde{I}_c * p_c)||_2^2 + \gamma||\tilde{I}_c||_2^2 \quad [7]$$

where $\tilde{I}$ is the unknown variable, p_c is the PSF p_λ integrated over the wavelength sensitivity κ_c in a narrow band around the RGB wavelengths, and $\gamma > 0$ is an (optimized) regularization parameter.

When the PSF discretization size matches the sensor pixel size, the pixel integration operator S matches the identity. Then, the optimization problem in Equation 7 can be solved in closed form, and the solution is given as Wiener filtering under the simplifying assumption of circular boundary conditions, which we can approximate by symmetric padding of y_c. the Wiener filtering operation is given by:

$$\tilde{I}_c = \mathcal{F}^{-1}\left\{\frac{\bar{p_c}^*}{|\bar{p_c}|^2+\gamma}\mathcal{F}\{y_c\}\right\} \quad [8]$$

where $\bar{p_c} = \mathcal{F}\{p_c\}$ is the optical transfer function, and multiplication and division are performed element-wise. When the pixel size is larger than the discretization of the PSF, we can use a fixed number of conjugate gradient steps to solve Equation 7.

B. End-to-end Optimization Framework. We develop a framework for optimizing the end-to-end pipeline in Pytorch. The optimization is conducted using the stochastic gradient descent method. Note that each stage of our model in Subsection A is expressed as a differentiable module. The only optimization

parameter in the pipeline is the optical height map, h. The list of hyperparameters include the optical element size, sensor pixel size, propagation distance z, and sensor read noise level.

We optimize our model on a dataset of RGB images. The optimization variables in the optical element and reconstruction method are optimized with respect to the mean-square error (MSE) loss over the dataset:

$$L(\tilde{I_c}, I_\lambda)_{MSE} = \sum_{c \in \{R,G,B\}} ||\tilde{I_c} - I_c|| \quad [9]$$

where $\tilde{I_c}$ denotes the reconstructed image, and I_λ denotes the ground truth image. To further encourage smoothness on the height map, we introduce Laplacian L1 regularize on the height map to encourage sparse solution:

$$R(H) = ||Laplacian(H)||_1 \quad [10]$$

Where H denotes the height map of the optical element. The final loss term used in our pipeline is a weighted combination of MSE loss and Laplacian L1 regularization:

$$L(\tilde{I_c}, I_\lambda, H)_{total} = L(\tilde{I_c}, I_\lambda)_{MSE} + \alpha R(H) \quad [11]$$

Where α is a scaling constant for regularizer. We set α in the scope of this report.

B.1. Optimization algorithms.. We experienced with several optimization algorithms extensively used in the deep learning community. We find that for this particular task, stochastic gradient descent (SGD) (9) performs significantly better than other dynamic optmization methods, such as AdamW. (10).

C. Feasibility Constraints of Fabrication. Developing the proposed optimization framework in a way that is compatible with real-world manufacturing constraints is of utmost importance. While it is possible to represent the optical element via a discretized height map h using a standard basis, there are known basis of representations that are compatible with the manufacturing requirements for various types of optical elements. Diffractive optical elements (DOEs), flat (thin) lenses that rely on small phase delays induced by ultra-small features etched in typically by a photolithography process, can be represented using a Fourier basis. Refractive optical elements, such as concave or convex lenses, can be represented using Zernike polynomials (11). By representing the height map h as a sum of weighted Fourier basis functions or Zernike polynomials up to a certain order, one can often directly plug them into manufacturing machines to produce the appropriate element. For the purpose of this report, we only utilize the Fourier basis representation that is appropriate for manufacturing a DOE. Finally, to address manufacturing tolerances, random uniform noise is added to the height map h during optimization.

3. Datasets and Implementation

A. Datasets. To train and evaluate our model, we use the DIV2K dataset (12). The datasets consists of RGB images with a large diversity of contents. The DIV2K dataset is further divided into train and validation sets, each of which consists of 800 and 100 samples, respectively. Notice that the images do not have corresponding depth maps, please refer to Subsection C on how we assign depth maps.

B. Data Pre-processing. To pre-process input data, we read input images and normalize pixel values to the range of [0, 1], since we're using deep learning frameworks that generally expect input tensors to be normalized. Next, to increase size of training data and promote generalization, we apply a random crop to sample an image patch of resolution 512×512. No further data augmentation is performed. We feed the aforementioned image patch to our end-to-end framework, and treat the same input image patch as the ground truth image.

C. Depth Map Generation. As mentioned previously, the DIV2K dataset does not contain depth maps for sample images. We need to arbitrarily assign depth maps for each image. To do that, we randomly select a depth from one of five depth options: 0.5 m, 0.67 m, 1 m, 2 m, and 1000 m. We use 1000 m as a proxy for optical infinity. We use 1000 m and optical infinity interchangeably in this paper. Then, the randomly sampled depth is applied to the entire image patch, effectively placing the two-dimensional planar image plan at that given depth. Such depth map assignment is consistent with the implementation described in the original paper (6).

4. Experiments

To evaluate our end-to-end model, we conduct two synthetic experiments, at increasing difficulty. For both experiments, the simulation parameters above are taken from the original paper (6), subject to memory constraints. Note that our optimization parameters differ from (6).

A. Simple Achromatic Lens. As an initial evaluation for our framework, we build a simple lens model that can focus light at optical infinity. *Simple* denotes that there is no deconvolutional filter involved and that the loss function (see Eq. (11)) is simply calculated between the simulated image output at the sensor level and the ground truth. *Achromatic* means that we're encouraging the optical element to be spectrally invariant, i.e., being able to focus light across multiple wavelengths.

A.1. Optimization Parameters. We use stochastic gradient descent (SGD) with momentum = 0.5. The initial learning rate is set to $5e-1$ and is maintained throughout training. We use a batch size of 8.

A.2. Simulation Parameters. We set aperture diameters to 5 mm, and sensor distance to 25 mm. We set the refractive indices of the optical elements to be 1.4648, 1.4599, and 1.4568 for each color channel, where the wavelength for each channel is 460 nm, 550 nm, 640 nm, respectively. We use a wave resolution of 2496×2496, and a sample interval of 2μm. Finally, we use a height tolerance of 20 nm. Finally, the Gaussian sensor read noise variance σ^2 is sampled from a uniform distribution between 0.1% and 2%.

Contrary to the process introduced in Section 3C, all depth maps in the simple achromatic lens experiment are assigned to 1000 m, since we are not encouraging this model to be depth-invariant. Moreover, in contrast to (6), we find that a standard basis representation for optimizing the height map of the optical element fails to converge to a solution. Instead, we explore utilizing a Fourier basis representation for optimization.

A.3. Results. When using a standard basis of representation of the height map, we fail to converge to a reasonable solution

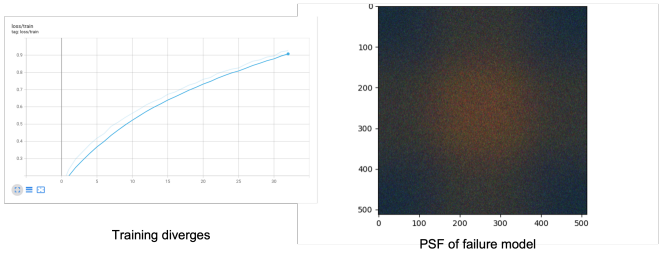


Fig. 2. Failure result of using a standard basis of representation of the height map element. When using a standard basis to represent the height map, the training loss diverges starting from epoch 1. As the result, the model fails to converge to a reasonable PSF. The insert shown on the right is the PSF at epoch 1, which mainly consists of noise and hardly any structured components.

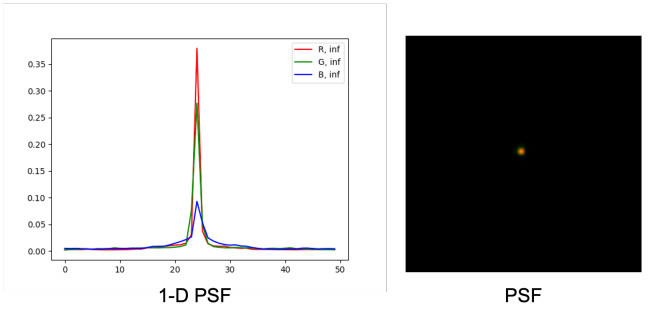


Fig. 3. Point Spread Function (PSF) of the optimized Simple Achromatic Lens. By using Fourier basis to represent the optical element, the model converges and the optimized model is able to focus light at infinity, evident in the compactness of one-dimensional PSF (left) and two-dimensional PSF (right). infinity, the model performs best in the red channel, and progressively worse in green and blue. Note that the height map has a low-pass filter, forcing frequencies above 50% to be zero.

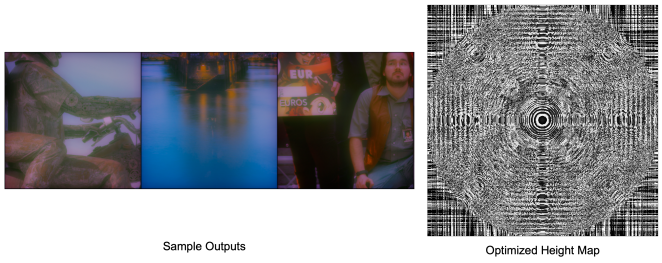


Fig. 4. Sample outputs from the optimized Simple Achromatic Lens. The optimized height map corresponding to the PSFs in Figure 3 is shown on the right. This height map produces output images shown on the left, where each image has a uniform depth map of optical infinity.

for a variety of optimization parameters. In fact, as shown in Figure 2, the training loss appears to diverge.

To combat this divergence, we suspect a representation with fewer high frequency components will help aide training convergence, effectively acting as an additional smoothness prior. We utilize a Fourier basis representation where we set 50% of the highest frequency Fourier coefficients to zero. This model successfully converges to an adequate solution, and the optimized PSF can be seen in Figure 3. Notice that at optical infinity, this model focuses light in the red wavelength the best, and then progressively worse and the wavelength deceases towards green and blue. This is consistent with the two dimensional PSF, where the red and green airy disk is more visible. The PSFs shown in Figure 3 is produced by

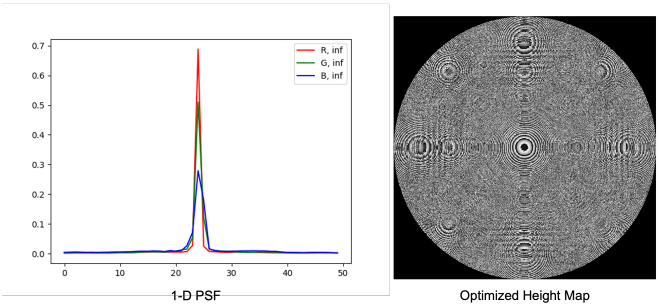


Fig. 5. Outcome of removing low pass filter on height map in the Achromatic Simple Lens model. By removing the low-pass filter on the height map and re-training the model, the model achieves better convergence. Notice that the model is now able to utilize high-frequency components to construct a sharp circular boundary (right). The resulting PSFs (left) have higher peaks than those from the model with a low-pass filter (shown in Figure 3 and 4).

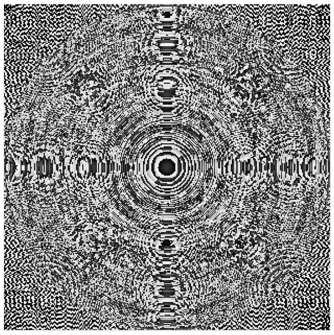
optimized height map element, whose height map is shown in Figure 4. Intuitively, the optimized height map resembles a profile that matches a Fresnel lens. Sample formed images by the optimized optical element are also shown in Figure 4. Note that all images are placed at optical infinity. Observe that our Achromatic Simple Lens model produces sharp images at this particular depth, however, a moderate amount of chromatic aberrations is present, since this model is comparatively worse at focusing light in the blue wavelength at optical infinity.

We postulate that this Fourier basis representation actually helps stabilize training irrespective of removing higher frequency components. To test this conjecture, we remove the low pass filter on the height map and allow all Fourier coefficients to have a non-zero value. Interestingly, easing this smoothness prior actually improves the performance of the model, as shown by the PSF in Figure 5. This suggests that the Fourier representation itself leads to faster, more stable training and better convergence.

B. Achromatic Extended Depth of Field (AEDoF). We present a second experiment that is simultaneously optimized for two objectives: (1) a PSF is spectrally invariant (achromatic), and (2) a PSF that is depth invariant (Extended Depth of Field). Notice that our set of objectives differ from that for the Simple Achromatic Lens mentioned in Subsection 4.A, and that lead to a number of main differences between the two models. First, we randomly assign an uniform depth map for each input image crop (see Subsection 3.C for details). Second, the AEDoF model now consists of a deconvolutional layer to perform image reconstruction. Next, we perform additional hyper-parameter tuning for the AEDoF model, which results in different optimization parameters. Additionally, some of the simulation parameters differ between the two models in the original paper (6), which is accounted for in our experiment.

B.1. Optimization Parameters. We use stochastic gradient descent (SGD) with momentum $= 0.5$. The initial learning rate is set to $5e-1$. We use a StepLR scheduler (13) with `step_size` $=$ 1 and gamma $= 0.8$ to decrease learning rate over time. We use a batch size of 4 due to memory constraint.

B.2. Simulation Parameters. We set aperture diameters to 5 mm, and sensor distance to 35.5 mm. We set the refractive indices of the optical elements to be 1.4648, 1.4599, and 1.4568 for each



Optimized Height Map

Fig. 6. The optimized Height map of the AEDoF model. The corresponding sample PSFs and sample outputs are shown in Figures 7 and 8.

color channel, where the wavelength for each channel is 460 nm, 550 nm, 640 nm, respectively. We use a wave resolution of 1248×1248, and a sample interval of 4μm. We use a Fourier basis representation and set the 37.5% highest frequencies to zero. A height tolerance of 20 nm is used. Finally, the Gaussian sensor read noise variance σ^2 is sampled from a uniform distribution between 0.1% and 2%.

B.3. Results. The optimized height map element for AEDoF is shown in Figure 6. Figure 7 shows the PSF of the AEDoF model. The 1-D PSFs are most compact at depth = 0.5 m, and then gets progressively worse as distance increases towards infinity. However, the PSFs are still moderately compact at optical infinity, suggesting that the PSF does not significantly alter with depth. This observation is consistent with the PSF stack shown the right of Figure 7, where the shape of the PSF does not change drastically between 0.5 m and infinity.

Figure 7 shows sample outputs of the same input when placed at different depths. Notice that at farther depths like 2 m and 1000 m, there exists purple fringing in the output, suggesting the existence of chromatic aberration. At closer depths, the output is marginally sharper and free of purple fringing. From visual inspection of the output, although the model performs the best at 0.5 m, the reconstructed quality does not drastically degrade as depth increases towards infinity, suggesting a degree of depth invariancy in our model. Further, since the chromatic aberration is minimal for a single-component optical system, we claim that our optimization results in a model that achieves AEDoF.

B.4. Ablation Study. In order to further evaluate the effectiveness of our optimization scheme, we train a model that is only optimized for focusing images placed at infinity. To do that, we set the depth map of all training image patches to be 1000 m. We call this model Achromatic Shallow Depth of Field (ASDoF). Note that the only difference between ASDoF and AEDoF (referenced in Subsection 4.B.3) is that AEDoF is only trained on images placed at infinity, while ASDoF is trained on images across all five depths. All other components and parameters (optimization and simulation) are the same between ASDoF and AEDoF. Comparing to Achromatic Simple Lens (Subsection 4.A), ASDoF has the sample objective (achromatic imaging at infinity), but ASDoF has some added components and different parameters, such as the deconvolution layer.

Figure 10 shows a comparison between outputs from AEDoF and ASDoF. Notice that while AEDoF produces relatively sharp images across all depth ranges, ASDoF only produces significantly better outputs at 1000m, but the image quality drastically degrades when at all other depth ranges. This suggests that our the AEDoF model is effectively optimized to be depth-invariant, while ASDoF is solely optimized for a singular depth at infinity. Next, observe that both models achieve achromaticity given the moderate amount of chromatic aberration. Further, the outputs from AEDoF have slight purple fringing (most visible at 1000 m), while those from ASDoF have slight green fringing (most visible at 0.67 m and 0.5 m). This indicates that the two models prioritize different channels. Specifically, the worst-performing channel for AEDoF and ASDoF are red and green, respectively. Notice that there is no mechanism in our training scheme that encourage the two model to prioritize different wavelengths, and we believe that randomness in initialization and subsequent gradient descent steps likely lead to different converges between the two models.

The corresponding one-dimensional PSFs of the two models are shown in Figure 9. Observe that the peak intensity for ASDoF (at optical infinity, maximum intensity = 0.25) is significantly better than PSFs for AEDoF for at any depth (maximum intensity = 0.16). However, at 1 m and 0.5 m, the PSFs of ASDoF have lower peaks than those from AEDoF. The two observations also indicate that AEDoF is significantly more depth invariant than ASDoF.

5. Discussion

In this report, we have implemented the end-to-end optics and image processing pipeline proposed by Sitzmann *et al.* (6). We successfully utilized this pipeline to learn the height map for a diffractive optical element and subsequent post-processing image reconstruction algorithm that can achieve achromatic imaging, and achromatic extended depth-of-field (EDoF) imaging.

Beyond the work shown in this report, the same end-to-end framework has been extended for a variety of applications. In the same work of Sitzmann *et al.*, the authors used the end-to-end framework to learn the optical element and image reconstruction algorithm for a super-resolution imaging system. Here, the optimized model achieves super-resolution by multiplexing sub-pixel-shifted image copies on the sensor, and using the known optimized PSF during image reconstruction. Wu *et al.* (14) learnt the optimal phase mask and reconstruction algorithm for single frame, single viewpoint, passive 3D imaging. A feature of this work, although proposed by earlier authors including Sitzmann *et al.*, is the use of a deep convolutional neural network for image reconstruction. Tseng *et al.* (15) optimize compound optics together with hardware and software image post-processors. To do this, they modify the PSF simulation proposed in Sitzmann *et al.* to a neural network-based model that has been trained predict spatially-varying PSFs and vignette given a set of optics parameters such as surface thicknesses, intervals, refractive indices, and surface parameters for every element. Finally, Hinojosa *et al.* (16) optimized an optical element and reconstruction algorithm to perform privacy-preserving human pose estimation. This work offers an interesting paradigm for the end-to-end optics and image processing framework – namely, that it optimizes an optical element to contain enough abberations as to preserve privacy while still being able to form an image that can be

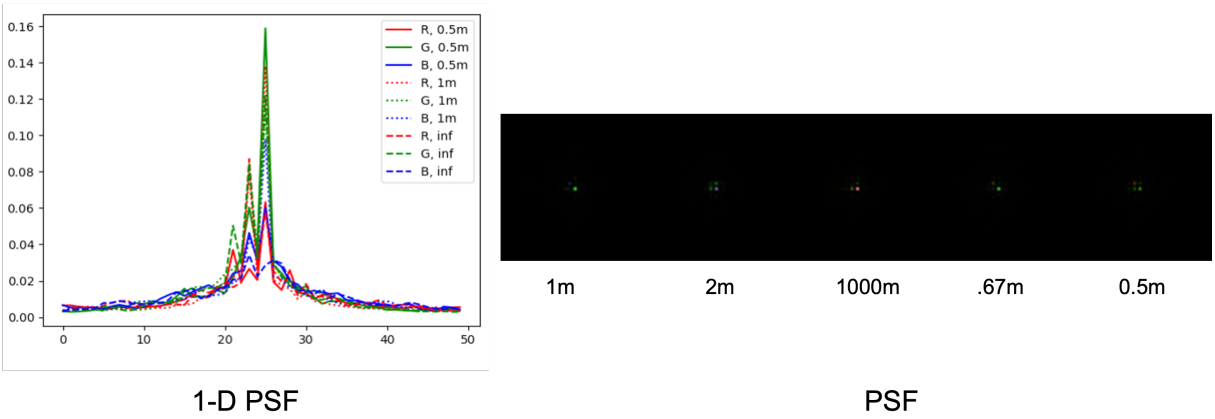


Fig. 7. The PSFs for the AEDoF model. The diagram on the left shows that the PSFs are most compact at $0.5\ \mathrm{m}$ and maintains the compactness as depth extends to infinity. The two-dimensional PSFs for each depth are shown on the right.

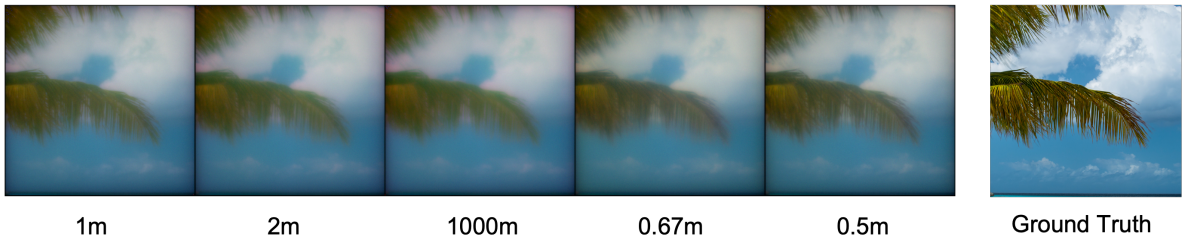


Fig. 8. Sample outputs from the AEDoF model. As the depth of the input image increases from $0.5\ \mathrm{m}$ to $1000\ \mathrm{m}$, the image quality does not drastically degrade, indicating that the model is fairly depth invariant. Although, as distance increases, a slight loss of sharpness and introduction of purple fringing is visible.

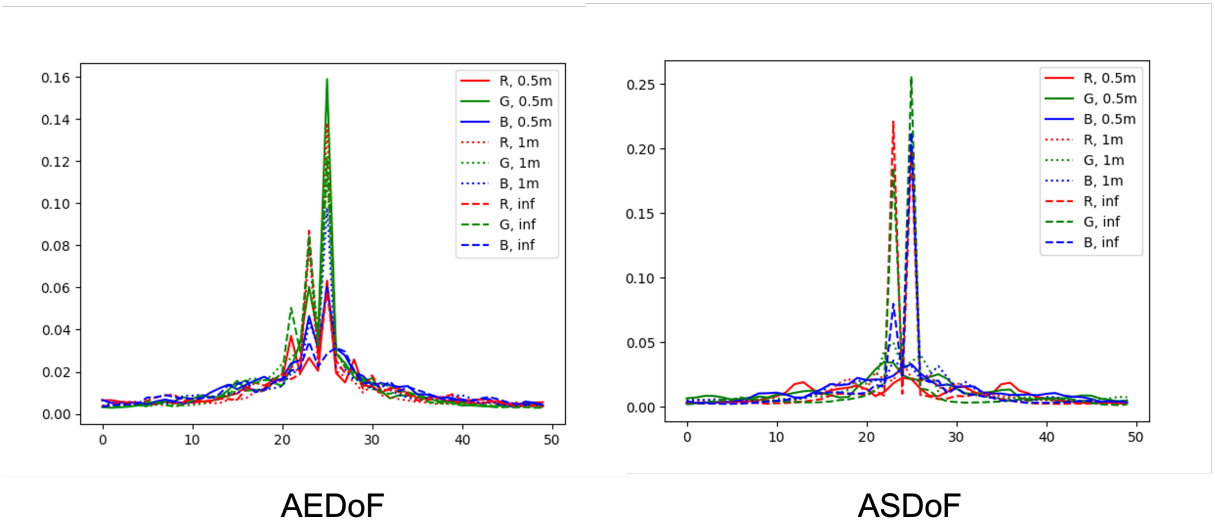


Fig. 9. A comparison of the PSFs of the AEDoF model and the ablation ASDoF model. While the PSFs at all three depths are compact for AEDoF, ASDoF has the most-compact PSF at infinity, but the PSFs flatten considerably at other depths. This indicates that AEDoF is optimized for extended depth of field, while ASDoF is only optimized for focusing at optical infinity.

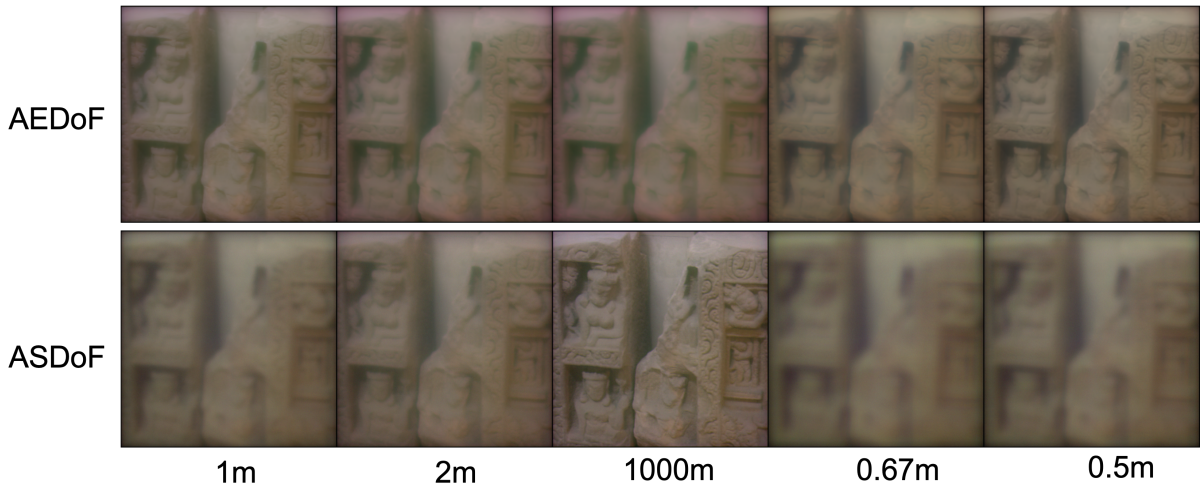


Fig. 10. Sample outputs from AEDoF and ASDoF models. AEDoF produces fairly sharp images across all depth ranges, with a slight loss of sharpness and addition of purple fringing towards $1000\ \mathrm{m}$. In contrast, ASDoF produces the sharpest result at $1000\ \mathrm{m}$, but the image quality quickly degrades as depth decreases.

used to predict human pose.

Overall, end-to-end optics and image processing offers an exciting framework to be utilized when designing application-specific vision systems. Today, such systems rely not on the raw measurement on the sensor, but the final post-processed result for a particular application. Therefore, by jointly optimizing both the optics and image processing pipeline, these systems can learn new methods of encoding visual cues that can then be decoded to reconstruct a final result. We hope that future works continue to modify and add to the differentiable optics model, for example by adding multiple sensors, exposure times, shutter speeds, or even novel optical elements such as metamaterials (17), which will enable new optimal vision systems to be learnt for applications beyond those discussed in this report.

1. Chen C, Chen Q, Xu J, Koltun V (2018) Learning to See in the Dark. pp. 3291–3300.
2. Tan RT (2008) Visibility in bad weather from a single image in *2008 IEEE Conference on Computer Vision and Pattern Recognition*. pp. 1–8. ISSN: 1063-6919.
3. Smith WJ (2007) *Modern Optical Engineering*. (McGraw Hill Professional), 4th edition. Google-Books-ID: DrtM_bAnf_YC.
4. Nayar SK (2006) Computational Cameras: Redefining the Image. *Computer* 39(08):30–38. Publisher: IEEE Computer Society.
5. Raskar R, Agrawal A, Tumblin J (2006) Coded exposure photography: motion deblurring using fluttered shutter in *ACM SIGGRAPH 2006 Papers*, SIGGRAPH '06. (Association for Computing Machinery, New York, NY, USA), pp. 795–804.
6. Sitzmann V, et al. (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):1–13.
7. Goodman JW (2005) *Introduction to Fourier Optics*. (Roberts and Company Publishers). Google-Books-ID: ow5xs_Rtt9AC.
8. Antipa N, Necula S, Ng R, Waller L (2016) Single-shot diffuser-encoded light field imaging in *2016 IEEE International Conference on Computational Photography (ICCP)*. pp. 1–11.
9. Sutskever I, Martens J, Dahl G, Hinton G (2013) On the importance of initialization and momentum in deep learning in *Proceedings of the 30th International Conference on Machine Learning*. (PMLR), pp. 1139–1147. ISSN: 1938-7228.
10. Loshchilov I, Hutter F (2019) Decoupled Weight Decay Regularization. *arXiv:1711.05101 [cs, math]*. arXiv: 1711.05101.
11. Born M, Wolf E (1999) *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*. (Cambridge University Press, Cambridge), 7 edition.
12. Agustsson E, Timofte R (2017) Ntire 2017 challenge on single image super-resolution: Dataset and study in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.
13. (2019) StepLR — PyTorch 1.10.0 documentation.
14. Wu Y, Boominathan V, Chen H, Sankaranarayanan A, Veeraraghavan A (2019) PhaseCam3D — Learning Phase Masks for Passive Single View Depth Estimation in *2019 IEEE International Conference on Computational Photography (ICCP)*. pp. 1–12. ISSN: 2472-7636.
15. Tseng E, et al. (2021) Differentiable Compound Optics and Processing Pipeline Optimization for End-to-end Camera Design. *ACM Transactions on Graphics* 40(2):1–19.
16. Hinojosa C, Niebles JC, Arguello H (2021) Learning Privacy-Preserving Optics for Human Pose Estimation. pp. 2573–2582.
17. Cai W, Shalaev VM (year?) *Optical metamaterials*. (Springer) Vol. 10.