On the use of differentiable optical models for lens and neural network co-design

Marius Dufraisse, Pauline Trouvé-Peloux, J-B Volatier, Frédéric Champagnat

On the use of differentiable optical models for lens and neural network co-design

M. Dufraisse[a], P. Trouvé-Peloux[a], J-B. Volatier[b], and F. Champagnat[a]

[a]DTIS, ONERA, Université Paris-Saclay, F-91123, Palaiseau, France
[b]DOTA, ONERA, Université Paris-Saclay, F-91123, Palaiseau, France

ABSTRACT

Co-design methods started to incorporate neural networks a few years ago when deep learning showed promising results in computer vision. This requires the computation of the point spread function (PSF) of an optical system as well as its gradients with respect to the optical parameters so that they can be optimized using gradient descent. In previous works, several approaches have been proposed to obtain the PSF, most notably using paraxial optics, Fourier optics or differential ray tracers. All these models have limitations and strengths regarding their ability to compute a precise PSF and their computational cost. We propose to compare them in a simple co-design task to discuss their relevance. We will discuss the computational cost of these methods as well as their applicability.

Keywords: Co-design, Neural network, Optics

1. INTRODUCTION

Imaging systems are composed of several components : most notably a lens, a sensor and a software processing. The interactions between these parts are tricky to model and understand. As such the most common way to design an imaging system is to have experts from various fields design separately each of the components as best as they can.

The process to design a lens starts with a design and optimizes it so that it is tuned for the desired application. The initial design can be from a book, patent database or a generative model.[1] In any case an experienced engineer is required to choose a design with relevant properties. Optical design software, such as Zemax or CodeV, have the capability to propagate rays through the lens, compute a merit function which evaluates the qualities of the lens, and optimize the optical parameters to improve this merit function. This workflow yield good results however the heuristics used to choose a merit function are often unknown and it can be difficult to come with a user defined merit function that take the processing into account.[2] Furthermore the optimization process needs to be guided by the optical engineer to reach a satisfactory system.

Image processing algorithms can perform many tasks ranging from image quality improvement to semantic analysis of images. Nowadays, most state of the art methods use deep neural networks. They need to be trained using many samples from real world data which might need to be annotated. Instead of gathering images for a specific lens design, neural networks are generally trained on existing datasets acquired using an off-the-shelf camera producing sharp images. This means they generally don't take into account the aberrations of a given lens.

The idea of co-design is to design better imaging systems by jointly optimizing the lens and the image processing. This method could yield processing algorithms that are better suited to exploit the aberrations of simpler lenses and thus help design optical systems with smaller optics. This approach proved to be useful for tasks such as extended depth of field[3] or depth estimation.[4] More recently some contributions were made regarding the use of neural networks in co-design.[5–8]

Jointly optimizing an optical system and a neural network using gradient descent requires the computation of the PSF of the optical system as well as its gradients with respect to the optical parameters. To do so, previous works have introduced several methods for PSF simulation. Each of them can get good results but most papers focus only on one of them and don't compare them. Here we compare the use of paraxial optics, Fourier optics and ray tracing. We discuss their advantages and analyze their behavior in a simple co-design task.

2. DIFFERENTIABLE OPTICAL SIMULATION

Four methods were proposed to perform differentiable PSF computation. We present them in this section.

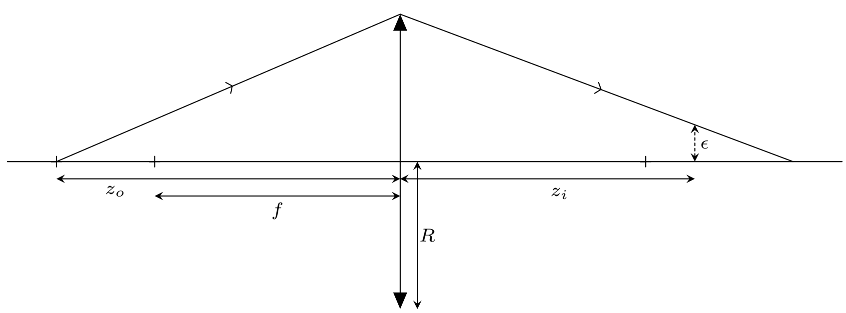
2.1 Paraxial optics

Paraxial optics is based on the thin lens model. For a lens composed of a single lens of focal length f and of radius R, it is possible to compute the radius ϵ of the image on a given plane of a point on the optical axis (see Figure 1). Under the Gaussian approximation we have

$$\epsilon = R\left|1 - z_i\left(\frac{1}{f} - \frac{1}{z_o}\right)\right|, \tag{1}$$

where z_i is the sensor distance and z_o the object distance. The PSF of the system can then be modeled as a normal distribution of variance $\rho\epsilon$, with ρ a constant determined arbitrarily in the literature.[4]

Figure 1. Relevant distances for the thin lens model.



When the system is made of several lenses this model remains useful but it is applied to an optical system with the same aperture and equivalent focal length that serves as a surrogate during the optimization. Those parameters are known for commercially available lenses. If a description of the surfaces of the lens is available they can be computed by tracing the path of a ray that enters the system parallel to the optical axis. This is enough to compute the equivalent focal length and the position of the second principal plane of the system.

The paraxial optics is a simplified model. It computes PSFs without taking into account aberrations except for axial chromatic aberrations. Furthermore it only has three parameters : the focal length, aperture and focus distance.

2.2 Fourier optics

Fourier optics[9] models the light as a complex amplitude wave to compute its propagation through an optical system. Fourier optics computes the phase shift induced by an optical system on the wave of light that passes through it.

In the paraxial approximation the amplitude and phase of a wave that goes through a thin lens is modified (at the point (x, y)) by a factor

$$A(x, y) \exp\left(-i\frac{\pi}{\lambda f}\left(x^2 + y^2\right)\right), \tag{2}$$

where A is the aperture function of the lens, λ the wavelength considered and f the focal length of the lens. For an object at distance z_o from the lens, the phase shift at coordinates (x, y) on a plane at z_i from the lens is

$$\mathcal{F}\left\{A(x,y)\exp\left(i\psi\frac{x^2+y^2}{R^2}\right)\right\}, \tag{3}$$

where R is the radius of the lens, $\mathcal{F}$ the Fourier transform and

$$= \frac{\pi R^2}{\lambda}\left(\frac{1}{z_o}+\frac{1}{z_i}-\frac{1}{f}\right). \tag{4}$$

Then the PSF is expressed as

$$PSF(x,y) = \left|\mathcal{F}\left\{A(x,y)\exp\left(i\psi\frac{x^2+y^2}{R^2}\right)\right\}\right|^2. \tag{5}$$

The value ofψ indicates the amount of defocus of the system.

This model also uses only three parameters that can be obtained as explained in Sec. 2.1. However Fourier optics can simulate diffraction and previous works have shown that this model can be used to co-design a phase mask and a neural network for extended depth of field, super-resolution[6,7] or depth estimation.[8] Aberrations could be introduced in this model by describing them in the Seidel or Zernike polynomial basis and adding them to the phase shift function.

2.3 Differentiable ray tracing

Differentiable Ray Tracing (DRT) models the light as rays that go through the system from a point source. The optical system is described as a sequence of surfaces. Each surface is defined as a geometrical shape (for instance a conic) described by parameters such as its curvature, its eccentricity or its distance to the next surface. The path of a ray is described by the intersection of the ray with every surface in the sequence. The trajectory of a ray is computed one surface at a time. Knowing the path of the ray until the surface n, its position after the surface $n+1$ can be found by first computing the intersection of the ray with the surface $n+1$ and then applying Snell law to compute how it is refracted. Computing the intersection of the ray and the surface can be tricky if a complex surface is used : for conics an analytic solution can be found but iterative methods are used for more complex surfaces. The paths of multiple rays coming from a point source with various directions can be computed to get a spot diagram of the lens. The PSF can then be obtained by integrating the energy of rays that reach each pixel. The PSF of an object away from the optical axis can also be computed with this method as it only requires to change the position of the point from where the ray start.

The method described above can compute a PSF, it relies on automatic differentiation[10] to get the gradient of the PSF with respect to the optical parameters.

Given an initial lens description, this model can optimize all of its parameters for sources at any distance on and off the optical axis. This method was used to design imaging systems with extended depth of field[11] or extended field of view.[12]

2.4 Gaussian approximation

DRT is a powerful way to compute the PSF of an optical system both on and off axis. This makes it an interesting choice for co-design. However it needs to propagate many rays to precisely compute the PSF and thus tend to be slower than other models. To get a simulation time more suitable to the optimization of an imaging system it is possible to assume that the PSF is a Gaussian distribution. In this case less rays are required to estimate the parameters (center and covariance) of the PSF resulting in a faster simulation. This approach was used to design an imaging system with an extended depth of field.[5]

3. COMPARISON OF DIFFERENTIABLE OPTICS MODELS

Previous works have introduced multiple models to compute the PSF of an optical system. However, these models are never compared directly and it is unclear which one should be used for co-design. We first compare how the physical description of light used in each model limits their applicability. Then we compare the results obtained with each one of the four models on a simple co-design problem : the optimization of a double Gauss lens and neural network system for extended depth of field.

3.1 Benefits and limitations of each optical model

The four models described in the above section can all compute an approximation of the PSF of an optical system. However they are derived using assumptions that simplify their description.

The four models are all differentiable and applicable to co-designing an optical system and a neural network. The paraxial model relies on many approximations regarding both the shape of the PSF and the number of relevant optical parameters : it can be used to dimension the main parameters of an optical system without taking into account aberrations and diffraction. Fourier optics introduces diffraction in its modeling of PSFs and can account for some aberrations using classical polynomial decompositions of the phase function.[13] Ray-tracing methods can optimize all the parameters of a lens for point sources located anywhere in the field. However these models are more computationally expensive : they need to compute and store the path of many rays whereas paraxial and Fourier optics apply simpler operations to a small set of parameters. DRT with a Gaussian approximation trades precision for computation time by reducing the number of rays and interpolating the PSF. Finally, ray-tracing method need a precise description of the optical system and thus an initial starting point which has to be chosen with insight.

Table 1 gather these comparisons.

Table 1. Applicability of the various differentiable optics methods.

	Paraxial	Fourier	DRT	Gaussian DRT
Aberrations	No	Yes	Yes	Yes
Off axis	No	No	Yes	Yes
Diffraction	No	Yes	No	No
Number of parameters	3	3	All	All
Number of rays required			Hundreds	Dozens

Figure 2 shows the PSF obtained on the optical axis with each model for a double Gauss lens of focal length 100mm.

3.2 Practical comparison on a simple example

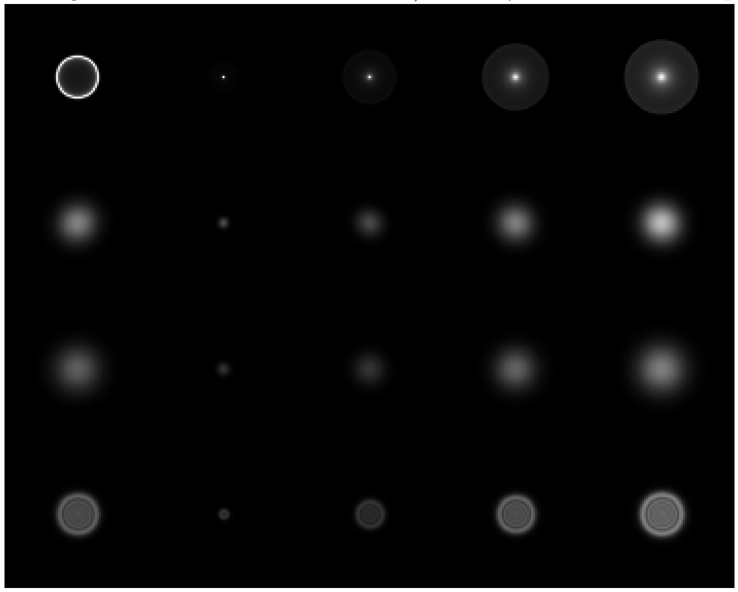
The various differentiable optical simulations model light differently and thus could lead to different results. As far as we know, no one compared them on a co-design task.

In order to analyze the choice of either paraxial, Fourier optics and Gaussian DRT models for lens and neural network co-design, we optimize optical parameters of a lens using these three models separately for the same use-case. DRT, that provides the most precise PSF simulation, is only used to simulate the actual lens PSF for these three settings, so that we can compare their final performance.

Extended depth-of-field is a problem often used to demonstrate the effectiveness of co-design. We set up a co-design task to produce the best possible images for objects at various distances. To keep all models relevant we only optimize the system for objects on the optical axis. We start the optimization with a Gauss doublet and a state of the art neural network for deconvolution. The sensor position is the only variable optical parameter: it sets which object distances will have the most defocus blur. The network processes the images to cancel this defocus blur and restores sharp images. The optimization is done in two steps:

1. The lens and the network are jointly optimized. During this step the PSF is simulated using either paraxial optics, Fourier optics or Gaussian DRT. This first step yields an optical system and a trained neural network for each model.

Figure 2. PSF obtained with each model (from top to bottom : DRT, gaussian DRT, paraxial, Fourier) for an object at various distances (from left to right : 10m, 20m, 30m, 40m, 50m). 256 rays were used to compute the PSF with DRT.



2. To evaluate the real-world performance of the optical system we compute its PSF using ray tracing to get a more precise estimation of the actual PSF of the physical system. Then the optical parameters are fixed to the value found during the first step and only the neural network is trained to fine-tune it and evaluate the real-world performance of the optical systems designed with each of the PSF simulation.

3.2.1 Optimization details

The lens design starts with a Gauss doublet of focal length 100mm (see Appendix A). The initial sensor plane is located so that the image of an object at 25m is focused. The sensor position is then updated using the Adam optimizer with a learning rate of 0.001. A Gaussian noise of variance 0.01 (for images in the range $[0, 1]$) is added after the optical simulation. The co-design is done for a monochromatic light at 530nm. The PSF is computed on an area of size 105μm divided into 21 pixels.

For the processing we use the RedNet[14] architecture. The Adam optimizer is used to minimize the L^1 loss of the imaging system. The learning rate is set to 0.001 and the batch size to 32. We train the model on patches of size 64×64 pixels in order to have a local processing that uses clues from optical aberrations to perform the restoration instead of relying on semantic information. The model is trained using 80% of the 22600 images from the Describable Textures Dataset;[15] the remaining 20% are used for validation.

3.2.2 Experimental results

The configurations obtained after training with each model are shown in Table 2. After the first optimization step, Fourier and paraxial optics are focused at roughly the same distance of around 35m whereas the Gaussian DRT model puts the focus distance closer to the center of the range at 27m. This shows that it is not necessary to take diffraction into account for this simple task as Fourier and paraxial optics converged to the same focus distance. The first step of the optimization process gets close to these final values fairly quickly (see Figure 4).

The focus distances obtained with each model correspond to compromises made between performances for objects at different distances from the lens. We use Gaussian DRT to estimate roughly the amount of defocus of the optimized systems : Figure 5 shows how the defocus blur changes with the object distance for the optical systems optimized using each of the models. This helps understand what is the trade-off made by the optimization. Fourier and paraxial optics focus distances are around 35m. The defocus for such an optical system

Table 2. Optimization results obtained with the different models. The first two columns show the validation loss reached when optimizing with each model as well as the loss obtained after fine-tuning the neural network with the ray-traced PSFs of the optimized optical systems. The last column provides the focus distance of each of the optimized systems.

Model	L^1 loss	L^1 loss with DRT PSF	Focus distance (m)
Paraxial	0.0415	0.0504	33.9
Fourier	0.0408	0.0509	35.7
Gaussian DRT	0.0429	0.0430	27.3

decreases with the position of the object. This configuration increases the depth of field at the cost of blurrier close objects. This results in a better average performance. On the other hand the Gaussian DRT model reaches a shorter focus distance. This yields an optical system with a larger defocus for far objects but smaller overall. This also results in less variations of the amount of defocus blur across the distance range which should be easier to process for the neural network. It can reach this configuration because its PSF is impacted by aberrations. In Figure 2 one can see that for objects after the focus plane the DRT PSF is more concentrated due to spherical aberrations. It means that Gaussian DRT can lead the optimization toward a configuration that better focuses close objects while having less defocus for far objects thanks to aberrations.

Figure 6 shows how the reconstruction error varies with the object distance. It compares the results of the imaging systems obtained after the two optimization steps. For a given optical system, sharper objects have a smaller loss and thus give better reconstructed images. However, the optical system obtained using Gaussian DRT have less defocus variation and as a result the processing can accommodate for both close and far objects. This yield an imaging system with a processing adapted to the entire range that can achieve better loss for all the object distances and significantly better loss between 20 and 30 meters. Note that the loss being lower for object at 10m compared to object at 15m comes from the fact that the PSF is computed for a 21×21 pixels area which is too small for the PSF of close objects. Regardless of this, the optical system optimized with Gaussian DRT remains better as the three systems would still be too blurry for close objects.

Table 2 provides the loss achieved by each model after the joint optimization of the lens and the neural network and when evaluated using ray-traced PSFs. After the first step of the optimization, Fourier have the best loss at 0.0408 whereas the paraxial and Gaussian DRT models reached 0.0415 and 0.0429 respectively. However, after freezing the optical system and optimizing only the neural network using fully ray traced PSFs, we can see that the performance of the systems optimized using Fourier or paraxial optics degrades more significantly (their loss increases by around 25%) than the ray-traced one (whose loss only increased by 1%). From all of this we can deduce that the two simpler models can be used for coarse adjustments of the focus distance but their results aren't optimal for an actual system. On the other hand, Gaussian DRT estimate the performance of an optical system more precisely and thus reaches better optical parameters.

4. CONCLUSION

We compared the models most often used in optics/neural-network co-design literature. They are based on different models of light propagation through a lens, thus each of them has its advantages and limitations. Fourier and paraxial optics are easier to simulate and can help solve simple problems such as determining the optimal focus distance for a lens described only by its aperture and focal length. However they don't work for off-axis objects and don't model aberrations. As a result they yield systems that aren't as good as what the Gaussian DRT model can produce. The Gaussian DRT model appears as a good trade-off between computation load and accuracy in a co-design perspective. It will be interesting to see if this two step optimization approach also work for off-axis objects or if it is necessary to use a fully ray-traced PSF computation for all the optimization. In this settings, the two simple models can still be useful to optimize the global parameters of the optical system to get a starting point more suitable for co-design before using ray-traced methods.

APPENDIX A. DETAILED INFORMATION ABOUT THE EXPERIMENTS

Our experiments were carried starting from the system described in the Zemax format in Table 3.

Figure 3. Evolution of the evaluation loss at each epoch of the training.

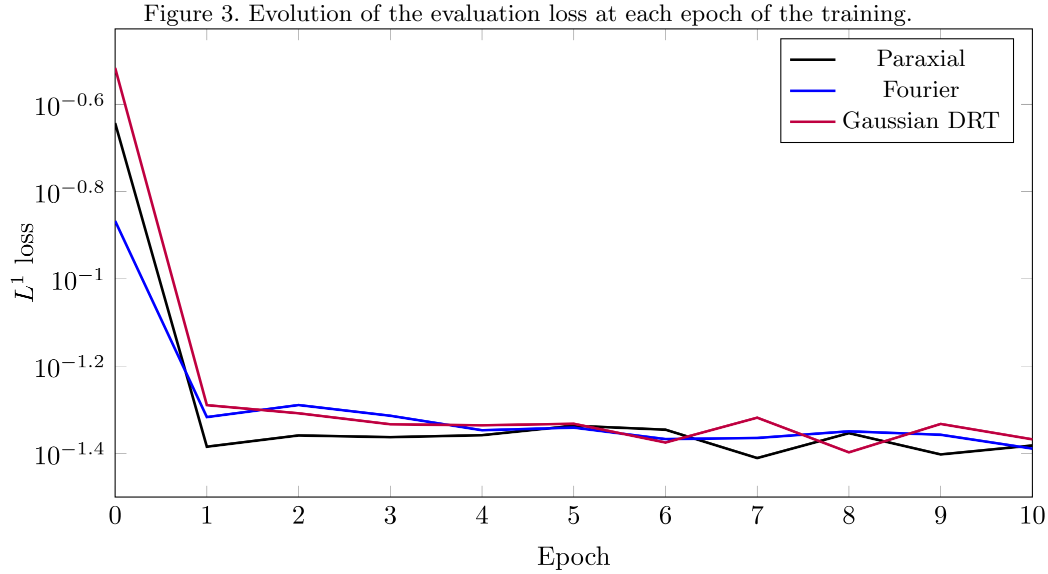


Figure 4. Evolution of the focus distance during training.

Figure 5. Size of the defocus blur as a function of the object distance for the lenses optimized with each model. The size of the defocus is computed using Gaussian DRT.

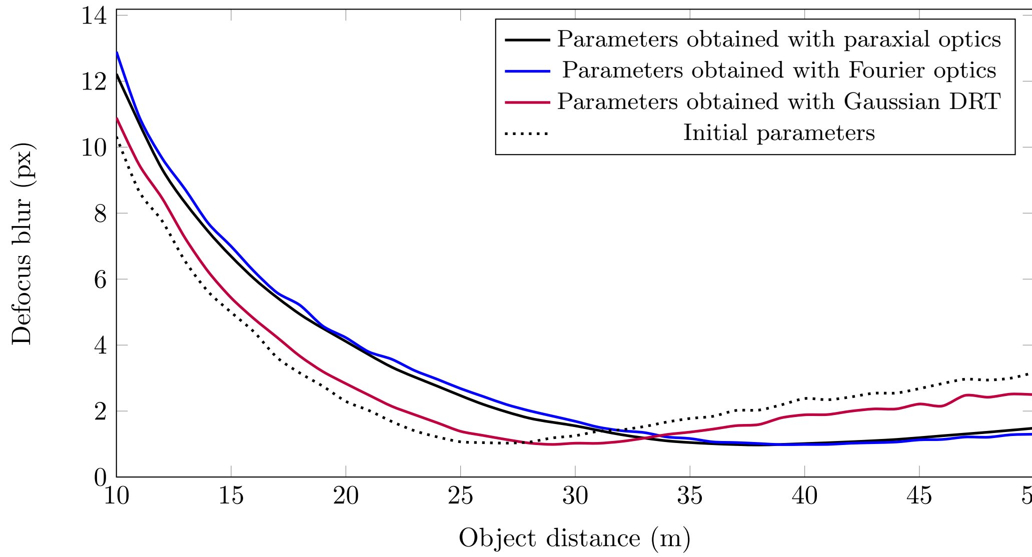


Figure 6. Variation of the loss with respect to the object distance after the second step of optimization of the neural network.

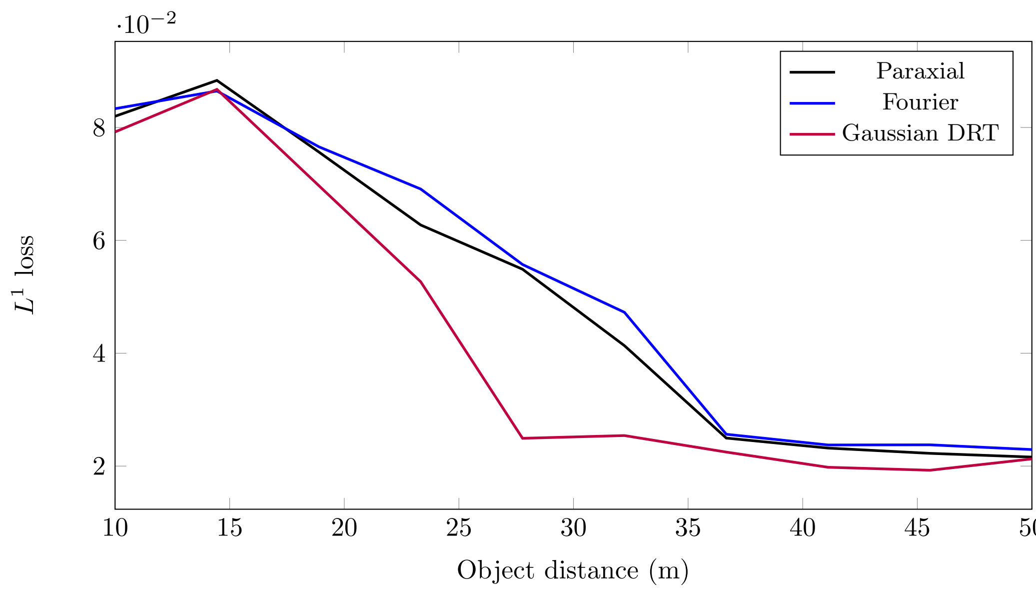


Table 3. The optical starting point for our experiment on co-design.

Surface	Curvature	Thickness	Material
Sphere 1	0.0184661976	8.74664	SK2
Sphere 2	0.0065564346	0.5	
Sphere 3	0.027815929	14.0	SK16
Plan 4		3.77696	F5
Sphere 5	0.04490361	14.253	
Aperture		12.4281	
Sphere 6	-0.03893318	3.777	F5
Plan 7		10.834	SK16
Sphere 8	-0.027041482	0.5	
Sphere 9	0.0050912	6.8581	SK16
Sphere 10	-0.014892576	58.5	
Image plan			

ACKNOWLEDGMENTS

This work was done during a PhD thesis co-funded by the ONERA and the Agence Innovation Défense.

REFERENCES

[1] Côté, G., Lalonde, J.-F., and Thibault, S., "On the use of deep learning for lens design," in [*OSA Optical Design and Fabrication 2021 (Flat Optics, Freeform, IODC, OFT)*], *OSA Optical Design and Fabrication 2021 (Flat Optics, Freeform, IODC, OFT)* , 120781A, Optica Publishing Group (2021).

[2] Fontbonne, A., *Conception conjointe combinaison optique / traitement : Une nouvelle approche de la conception optique de haut niveau*, theses, Université Paris-Saclay (Dec. 2021).

[3] Diaz, F., Goudail, F., Loiseaux, B., and Huignard, J.-P., "Increase in depth of field taking into account deconvolution by optimization of pupil mask," *Opt. Lett.* **34**, 2970–2972 (Oct 2009).

[4] Trouvé, P., Champagnat, F., Besnerais, G. L., Druart, G., and Idier, J., "Design of a chromatic 3d camera with an end-to-end performance model approach," in [*2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops*], 953–960 (2013).

[5] Hale, A. D., Trouvé-Peloux, P., and Volatier, J.-B., "End-to-end sensor and neural network design using differential ray tracing.," *Optics express* **29 21**, 34748–34761 (2021).

[6] Elmalem, S., Giryes, R., and Marom, E., "Learned phase coded aperture for the benefit of depth of field extension," *Opt. Express* **26**, 15316–15331 (Jun 2018).

[7] Sitzmann, V., Diamond, S., Peng, Y., Dun, X., Boyd, S., Heidrich, W., Heide, F., and Wetzstein, G., "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), 114 (2018).

[8] Haim, H., Elmalem, S., Giryes, R., Bronstein, A. M., and Marom, E., "Depth estimation from a single image using deep learned phase coded mask," *IEEE Transactions on Computational Imaging* **4**, 298–310 (Sep. 2018).

[9] Goodman, J., [*Introduction to Fourier Optics*], W. H. Freeman (2017).

[10] Griewank, A. et al., "On automatic differentiation," *Mathematical Programming: recent developments and applications* **6**(6), 83–107 (1989).

[11] Sun, Q., Wang, C., Qiang, F., Xiong, D., and Wolfgang, H., "End-to-end complex lens design with differentiable ray tracing," *ACM Transactions on Graphics (TOG)* **40**(4) (2021).

[12] Li, Z., Hou, Q., Wang, Z., Tan, F., Liu, J., and Zhang, W., "End-to-end learned single lens design using fast differentiable ray tracing," *Opt. Lett.* **46**, 5453–5456 (Nov 2021).

[13] Chang, J. and Wetzstein, G., "Deep optics for monocular depth estimation and 3d object detection," in [*Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*], (October 2019).

[14] Jiang, J., Zheng, L., Luo, F., and Zhang, Z., "Rednet: Residual encoder-decoder network for indoor rgb-d semantic segmentation," *arXiv preprint arXiv:1806.01054* (2018).

[15] Cimpoi, M., Maji, S., Kokkinos, I., Mohamed, S., , and Vedaldi, A., "Describing textures in the wild," in [*Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*], (2014).