Differentiable Strong Lensing for Complex Lens Modelling

Luca Biggio* ETH Zürich
Aymeric Galan EPFL
Austin Peel EPFL
Georgios Vernardos EPFL
Frédéric Courbin EPFL

Abstract

Strong lensing is a stunning physics phenomenon through which the light emitted from a distant cosmological source is distorted by the gravitational field of a foreground object distributed along the line of sight. Strong lensing observations are important, since, from their analysis, it is possible to infer properties of both the light-emitting source and the lens. In particular, precise lens modelling allows for the extraction of precious information on the distribution of dark matter in galaxies and clusters, which can provide tight constraints on several cosmological parameters. In this work, we consider the case where a comprehensive closed-form parametric model of the lens potential is only partially available, and we propose to model missing mass along the line-of-sight with a deep neural network. We incorporate the network within a fully differentiable, physically sound strong lensing simulator, and we train it via maximum likelihood estimation in an end-to-end fashion. Our experiments show that the model is able to effectively interact with the other components of the simulator and can successfully retrieve the underlying potential without any assumption on its form.

1 Introduction

In strong lensing, the light emitted from a distant cosmological source (e.g. a quasar or galaxy) is deflected by the gravitational potential exerted by a foreground object (e.g. an additional galaxy or cluster of galaxies) along the line of sight. This deflection leads us to observe an image that is a distorted version of what we would observe if the lensing object were not present [1]. Observations of strong lensing systems represent precious probes of the structure of the lens object, namely their dark matter content and its spatial distribution (see [2] and references therein). They can also be used to infer the properties of very distant cosmological objects owing to the fact that the lens magnifies the source light, allowing us to see objects that would otherwise be too faint [3].

Several efforts [4, 5, 6] have been made to use machine learning tools to extract information from lensing images, including both the parameters of the lens and of the source. Typically, convolutional neural networks (CNNs) [7] are used for this purpose, given their ability to process grid-like data structures. This class of techniques has resulted in dramatic gains in inference efficiency compared to classic Markov-chain Monte Carlo methods. Their main downside is that they need significantly large datasets for their training procedure. This often causes a performance degradation when the same network is tested on real data, which present inevitable discrepancies compared to simulated ones.

A different line of research leveraging the power of modern automatic differentiation (AD) tools [8] has recently emerged [9, 10]. In these works, the use of AD enables the optimization of parameters of the lensing system via gradient descent, and paves the way to gradient-based posterior sampling techniques such as Hamiltonian Monte Carlo [11].

In this work, we build on the work of [9, 10] and [12, 13], and we propose to infer the form of unknown perturbations to the lensing potential with a neural network in model-agnostic fashion. We

*Corresponding author: luca.biggio@inf.ethz.ch

Fourth Workshop on Machine Learning and the Physical Sciences (NeurIPS 2021).

do so by building a parameteric model combining physically principled and black box components in a unique differentiable strong lensing simulator. We introduce the basic principles of strong lensing along with our neural network-based parametrization of potential perturbations in Section 2. In Section 3, we present our preliminary results on two relatively simple lensing systems and in Section 4 we summarize our conclusions and we discuss future research directions.

2 Differentiable Strong Lensing

In the formalization of strong lensing, one usually starts from the so-called *thin lens approximation*, which assumes that all the mass is contained into a single plane. This is justified, since the dimensions of the lensing object are much smaller than the distances, D_{LS} and D_L (see Fig. 1), separating the lens from the source and the lens from the Earth, respectively. Under this approximation, the following equation relating the source-plane coordinates, $\vec{x}$, and the image-plane coordinates, $\vec{\xi} = (\xi_x, \xi_y)$, holds[2]

$$\vec{x} = \vec{\xi} - \vec{\alpha}(\vec{\xi}), \tag{1}$$

where $\vec{\alpha}$ is the deflection angle determined by the projected gravitational potential Ψ of the lensing object. This lensing potential can be written as an integral along the line of sight of the 3D potential ψ as

$$\Psi(\vec{\xi}) \equiv \int_{-\infty}^{\infty} \psi\left(\xi_x, \xi_y, z\right) \mathrm{d}z, \tag{2}$$

where z represents the coordinate orthogonal to the planes. From general relativity considerations, it can be shown that the deflection angle is related to the projected potential by the following equation

$$\vec{\alpha}(\vec{\xi}) = \frac{D_{LS}}{D_S} \frac{2}{c^2} \frac{\vec{\nabla}_{\vec{\xi}} \Psi(\vec{\xi})}{D_L}. \tag{3}$$

It is common to split the potential into two terms: one describing a macroscopic smooth component Ψ^M, which, given the source light, reproduces the observed multiple images, and a "substructure" component Ψ^s_θ which describes local effects due to any additional mass along the line of sight, either internal or external to the main lensing potential. In this work, we use a neural network with parameters θ to approximate Ψ^s_θ. We define the substructure component on a fixed coordinate grid and we train the network to map input locations in the grid to the corresponding pixel values. The final potential is then given by

$$\Psi = \Psi^M + \Psi^s_\theta. \tag{4}$$

The last step needed to generate simulated data is to apply the distortion introduced by the potential to the lens light intensity I_{src}. This can be done by exploiting the fact that lensing conserves surface brightness, and thus the image of the system in the lens plane, I_{lens}, is simply obtained by evaluating the source light distribution, I_{src}, on the lens plane. Hence, we have

$$I_{\mathrm{lens}}(\vec{\xi}) = I_{\mathrm{src}}(\vec{\xi} - \vec{\alpha}(\vec{\xi})). \tag{5}$$

Note that $I_{\mathrm{lens}}(\vec{\xi})$ will depend on the parameters θ of the network through the deflection angle $\vec{\alpha}$, which, according to Eq. 3, depends on the lens potential defined by Eq. 4. Therefore, in order to find the parameters θ, we exploit the fact that our entire pipeline is fully differentiable and allows us to calculate gradients of a likelihood function measuring the distance between observed data and the generated image I_{lens}. Our simulator is written in JAX [14] and our network is implemented in Flax [15], allowing us to exploit the AD engines at the core of these tools.

[2]The notation used in [10] is followed hereafter.

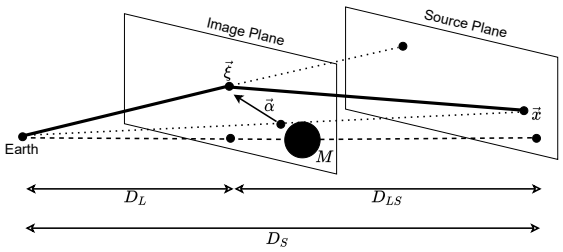

Figure 1: Illustration of the geometry of a strong lensing system under the thin lens approximation.

3 Experiments

We test our method on two different types of perturbations: a singular isothermal ellipsoid (SIE) and an external shear. The source light is modelled as a Sérsic ellipse, whereas the macroscopic potential is described by an SIE. More details on the form of these models and on the ground truth parameters characterizing all the components of the lensing system, including the ground truth perturbations, are reported in the Appendix. Generated images further include instrumental and atmospheric effects to mimic real data. More specifically, after generating the image of the lensed source, we convolve it with a gaussian point spread function and add random gaussian noise to account for measurement errors. For both experiments, we use the same model consisting of a simple fully-connected neural network with three hidden layers, each with 200 neurons. We model the full potential on a pixellated grid and, similarly to [16], the network takes as input a pixel entry in the image plane and outputs the corresponding intensity of the perturbation. Contrarily to [16], though, we use the Swish activation function [17], since we noted that other choices (e.g. ReLU and sine) led to instabilities in the calculation of the gradients of the final potential with respect to the input locations. In our experiments we found it beneficial to interrupt the training of the network after 200 epochs, since longer training resulted in overfitting the observational noise (see Appendix). The Adam optimizer [18] with learning rate 0.001 was used to minimize the mean squared error between simulated image and data.

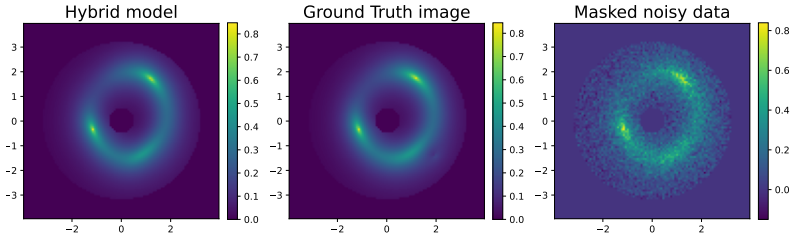

Figure 2: Results on the SIE perturbation. (Left) Output image from our hybrid model. (Middle) Ground truth denoised lensed image. (Right) Noisy image data.

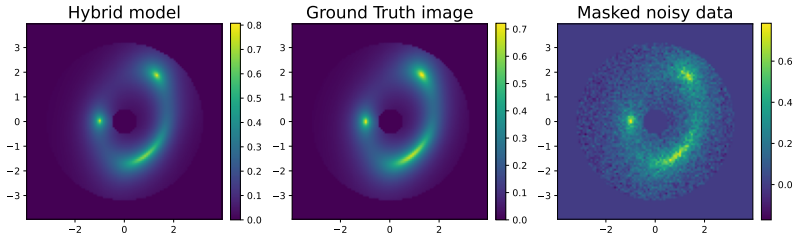

Figure 3: Results on the shear perturbation. (Left) Output image from our hybrid model. (Middle) Ground truth denoised lensed image. (Right) Noisy image data.

As shown in Figs. 2 and 3, the final lensed images generated by our hybrid simulators match the ground truth denoised images very closely. This performance results from the ability of the neural network to infer the shape of the perturbation given noisy data (see Figs. 4 and 5 for the SIE and the shear perturbation respectively).

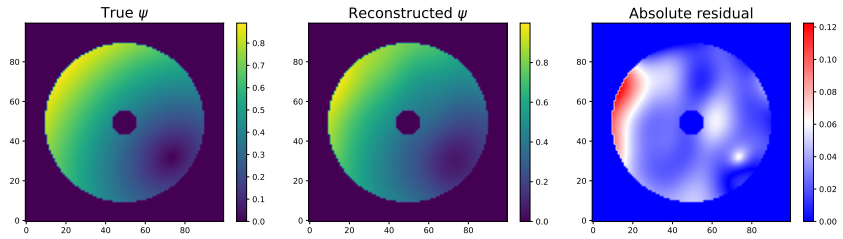

Figure 4: Results on the SIE perturbation. (Left) Ground truth perturbation. (Middle) Predicted perturbation. (Right) Absolute residual between prediction and ground truth.

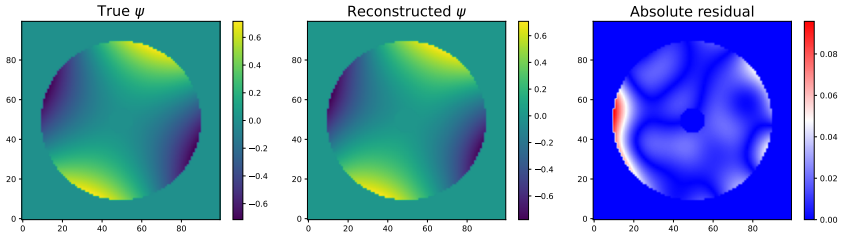

Figure 5: Results on the shear perturbation. (Left) Ground truth perturbation. (Middle) Predicted perturbation. (Right) Absolute residual between prediction and ground truth.

As a sanity check, we also assessed the performance of our network on fitting the perturbation with noise-free data. As shown in the Appendix, the model retrieves the ground truth potential with a very high level of accuracy.

4 Conclusions

In this work, we have proposed a new fully differentiable strong lensing simulator employing neural networks to learn unknown physics components in the lensing potential. We have shown that the method is effective in inferring the shape of both an SIE and shear perturbations, resulting in final simulated images in good agreement with observed data. Our method presents a number of interesting properties: first, it uses neural networks as black-box models only to learn unknown parts in the modelled systems; all the remaining components are either fixed or constrained by physically motivated parametric assumptions. Second, the network acts within a physically sound simulator, satisfying the well-known mechanisms underlying the strong lensing effect. Third, once trained, the network can interpolate between the pixels in the image plane and predict the potential values on points it did not use as training inputs. On the other hand, the current implementation of our method lends itself to a number of extensions and presents some limitations. Most importantly, due to its high expressivity, the network might overfit the training data (see Appendix for an example of this behaviour) by learning the noise, resulting in unrealistic potentials. Future work should focus on introducing regularization terms on the loss function penalizing physically unreasonable solutions. It will also be important to test the method on more complex and unstructured perturbations that might result in more subtle distortions in the final image. Additionally, we plan to extend this approach in such a way that a similar neural network to that used in our experiments can be included at any level of the strong lensing systems (e.g. to enhance the source model). We conclude by highlighting that the choice of a neural network as a black box function approximator is not the only one and that the power of our method mainly resides in its fully differentiable nature, making gradient-based optimization on the model parameters possible.

References

[1] Tommaso Treu. Strong lensing by galaxies. Annual Review of Astronomy and Astrophysics, 48(1):87–125, 2010.

[2] R. B. Metcalf, M. Meneghetti, C. Avestruz, F. Bellagamba, C. R. Bom, E. Bertin, R. Cabanac, F. Courbin, A. Davies, E. Decencière, and et al. The strong gravitational lens finding challenge. Astronomy Astrophysics, 625:A119, May 2019.

[3] Philip J. Marshall, Tommaso Treu, Jason Melbourne, Raphaël Gavazzi, Kevin Bundy, S. Mark Ammons, Adam S. Bolton, Scott Burles, James E. Larkin, David Le Mignant, David C. Koo, Léon V. E. Koopmans, Claire E. Max, Leonidas A. Moustakas, Eric Steinbring, and Shelley A. Wright. Superresolving Distant Galaxies with Gravitational Telescopes: Keck Laser Guide Star Adaptive Optics and Hubble Space Telescope Imaging of the Lens System SDSS J0737+3216. , 671(2):1196–1211, December 2007.

[4] Yashar D. Hezaveh, Laurence Perreault Levasseur, and Philip J. Marshall. Fast automated analysis of strong gravitational lenses with convolutional neural networks. Nature, 548(7669):555–557, Aug 2017.

[5] Laurence Perreault Levasseur, Yashar D Hezaveh, and Risa H Wechsler. Uncertainties in parameters estimated with neural networks: Application to strong gravitational lensing. The Astrophysical Journal Letters, 850(1):L7, 2017.

[6] Warren R. Morningstar, Yashar D. Hezaveh, Laurence Perreault Levasseur, Roger D. Blandford, Philip J. Marshall, Patrick Putzky, and Risa H. Wechsler. Analyzing interferometric observations of strong gravitational lenses with recurrent and convolutional neural networks, 2018.

[7] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. Neural Computation, 1(4):541–551, 1989.

[8] Atilim Gunes Baydin, Barak A. Pearlmutter, Alexey Andreyevich Radul, and Jeffrey Mark Siskind. Automatic differentiation in machine learning: a survey, 2018.

[9] Marco Chianese, Adam Coogan, Paul Hofma, Sydney Otten, and Christoph Weniger. Differentiable strong lensing: uniting gravity and neural nets through differentiable probabilistic programming. Monthly Notices of the Royal Astronomical Society, 496(1):381–393, May 2020.

[10] Konstantin Karchev, Adam Coogan, and Christoph Weniger. Strong-lensing source reconstruction with variationally optimised gaussian processes, 2021.

[11] Handbook of markov chain monte carlo. May 2011.

[12] Simon Birrer, Adam Amara, and Alexandre Refregier. Lensing substructure quantification in RXJ1131-1231: a 2 keV lower bound on dark matter thermal relic mass. Journal of Cosmology and Astroparticle Physics, 2017(05):037–037, may 2017.

[13] D. Bayer, S. Chatterjee, L. V. E. Koopmans, S. Vegetti, J. P. McKean, T. Treu, and C. D. Fassnacht. Observational constraints on the sub-galactic matter-power spectrum from galaxy-galaxy strong gravitational lensing, 2018.

[14] James Bradbury, Roy Frostig, Peter Hawkins, Matthew James Johnson, Chris Leary, Dougal Maclaurin, George Necula, Adam Paszke, Jake VanderPlas, Skye Wanderman-Milne, and Qiao Zhang. JAX: composable transformations of Python+NumPy programs, 2018.

[15] Jonathan Heek, Anselm Levskaya, Avital Oliver, Marvin Ritter, Bertrand Rondepierre, Andreas Steiner, and Marc van Zee. Flax: A neural network library and ecosystem for JAX, 2020.

[16] Vincent Sitzmann, Julien N.P. Martel, Alexander W. Bergman, David B. Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. In Proc. NeurIPS, 2020.

[17] Prajit Ramachandran, Barret Zoph, and Quoc V. Le. Searching for activation functions, 2017.

[18] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017.

[19] J. L. Sérsic. Influence of the atmospheric and instrumental dispersion on the brightness distribution in a galaxy. Boletin de la Asociacion Argentina de Astronomia La Plata Argentina, 6:41–43, February 1963.

[20] R. Kormann, P. Schneider, and Matthias Bartelmann. Isothermal elliptical gravitational lens models. Astronomy and Astrophysics, 284:285–299, 03 1994.

[21] Y. Yao, Lorenzo Rosasco, and Andrea Caponnetto. On early stopping in gradient descent learning. Constructive Approximation, 26:289–315, 2007.

A Checklist

1. For all authors...
   - Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [YES]
   - Did you describe the limitations of your work? [YES]
   - Did you discuss any potential negative societal impacts of your work? [N/A]
   - Have you read the ethics review guidelines and ensured that your paper conforms to them? [YES]
2. If you are including theoretical results...
   - Did you state the full set of assumptions of all theoretical results? [N/A]
   - Did you include complete proofs of all theoretical results? [N/A]
3. If you ran experiments...
   - Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [NO]. **This paper presents only preliminary results and it is part of a larger project close to its completion. Code and documentation for our library will be published soon.**
   - Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [YES]. **Details on network architecture, data structures and optimizers are included both in the main paper and in the Appendix.**
   - Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [N/A]
   - Did you include the total amount of compute and the type of resources used (e.g., 518 type of GPUs, internal cluster, or cloud provider)? [YES]. **Details are reported in the Appendix.**
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   - If your work uses existing assets, did you cite the creators? [N/A]
   - Did you include any new assets either in the supplemental material or as a URL? [NO]
   - Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   - Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   - Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   - Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   - Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

B Appendix

B.1 Experimental details

Our unperturbed lensing system is described by an elliptical Sérsic source light profile [19] and an SIE mass profile [20]. The first can be fully described by 7 parameters: the surface brightness I_0, the semi-major axis half light radius $R_{\rm sersic}$, the Sérsic index $n_{\rm sersic}$, the two eccentricity components e_1^s and e_2^s and the coordinates of the centre x_0^s and y_0^s. An SIE mass profile is described by 5 parameters, namely the Einstein radius $\theta_{\rm E}$, the eccentricity components e_1^l and e_2^l and the coordinates of the centre x_0^l and y_0^l. Our choices for these 12 parameters are listed in Table 1.
Finally, a shear potential is used to model additional angular structure in the final lensed image and results from unstructured additional matter distribution either within the cluster or distributed along the line of sight. It is parameterized by the two shear components, γ_1 and γ_2. Our choices for the parameters of the perturbations are listed in Table 2.
For all our experiments, we select a pixel size value equal to 0.08 arcsec (similar to data obtained with the Hubble Space Telescope) and each image is of size 100×100 pixels. All our experiments have been performed on the CPU of an Intel(R) Xeon(R) Gold 5120 processor.

Parameter Name	**Value**
I_0	10.0
$R_{\rm sersic}$	1.2 arcsec
$n_{\rm sersic}$	1.5
e_1^s	-0.09
e_2^s	0.03
x_0^s	0.4 arcsec
y_0^s	0.15 arcsec
$\theta_{\rm E}$	1.6 arcsec
e_1^l	0.15
e_2^l	-0.04
x_0^l	0.0
y_0^l	0.0

Table 1: Parameter choices of source and lens models. The source light profile amplitude is given in arbitrary units, as it represents a simple scaling of pixel values.

Parameter Name	**Value**
$\theta_{\rm E}^p$	0.15 arcsec
e_1^{lp}	0.0
e_2^{lp}	0.3
x_0^{lp}	1.9
y_0^{lp}	-1.4
γ_1	-0.1
γ_2	-0.1

Table 2: Parameter choices of the lens perturbations.

B.2 Results without noise

In this section, we repeat the experiment with the SIE perturbation with the difference that we remove the noise from the data. As shown in the figure below, the network infers very well the ground truth potential. The final image is indistinguishable from the true one, and this can be understood by looking at the gradient components of the perturbation, which match the truth almost perfectly.

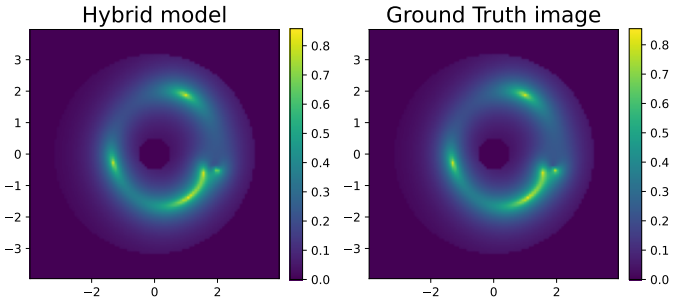

Figure 6: Results on the SIE perturbation with noiseless observations. (Left) Output image from our hybrid model. (Right) Noise-free image data.

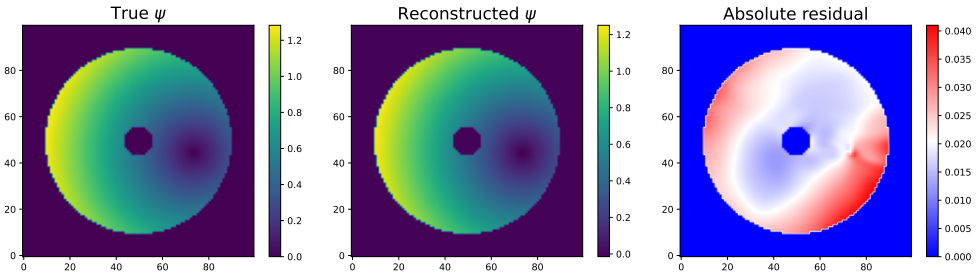

Figure 7: Results on the SIE perturbation with noiseless observations. (Left) Ground truth perturbation. (Middle) Predicted perturbation. (Right) Absolute residual between prediction and ground truth.

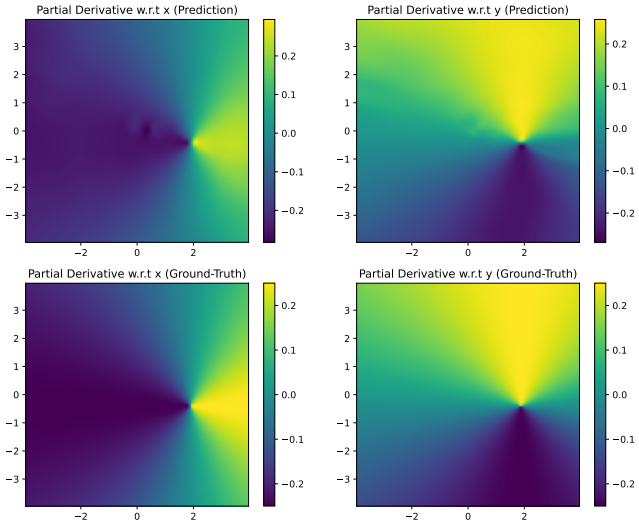

Figure 8: Results on the SIE perturbation with noiseless observations. The top (bottom) row shows the partial derivatives of the predicted (ground-truth) perturbation.

B.3 The risks of Overfitting

In this section, we investigate the negative effects of overfitting on the final reconstructed images. Overfitting results from the network learning the noise in the image. Looking at the figures below,

this effect manifests itself in the appearance of multiple artifacts in the final lensed image as well as in the potential and in its gradients. As shown in the paper, we empirically observed that this effect can be mitigated by applying an early stopping criterion [21] during training. As discussed in the conclusions, future work will focus on introducing meaningful regularization terms in the loss function in order to constrain the network to output more physically sound predictions.

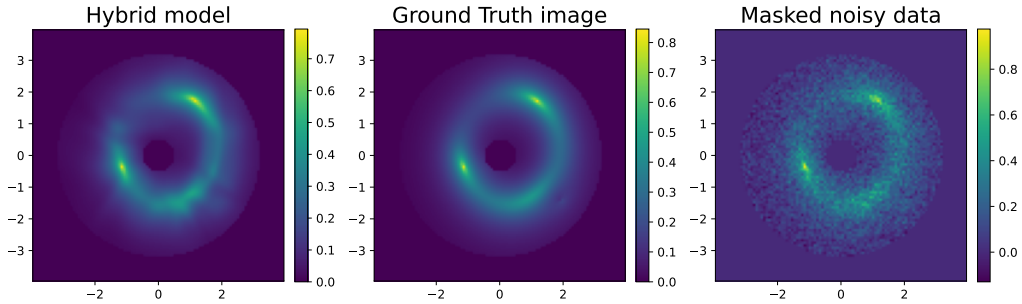

Figure 9: Results on the SIE perturbation with the network in the overfitting regime. (Left) Output image from our hybrid model. Note the numerous artifacts generated by the network. (Middle) Ground truth noiseless image. (Right) Noisy image data.

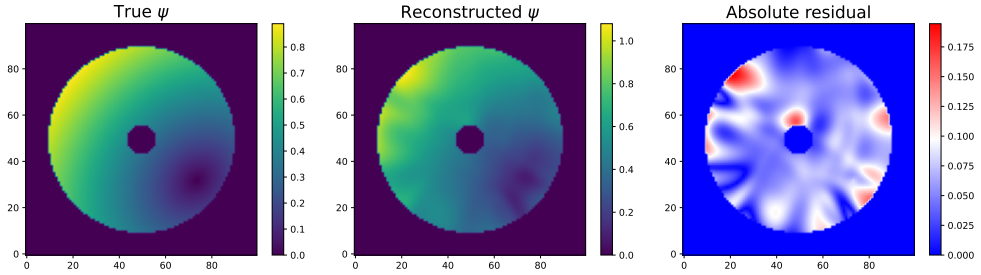

Figure 10: Results on the SIE perturbation with the network in the overfitting regime. (Left) Ground truth perturbation. (Middle) Predicted perturbation. Note the numerous artifacts generated by the network. (Right) Absolute residual between prediction and ground truth

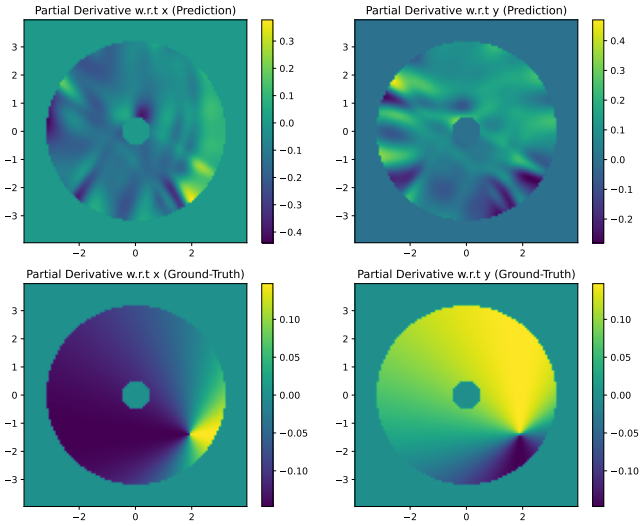

Figure 11: Results on the SIE perturbation with the network in the overfitting regime. The top (bottom) row shows the partial derivatives of the predicted (ground-truth) perturbation.