

---

# Simulation-Based Inference of Strong Gravitational Lensing Parameters

---

Ronan Legin\*<sup>1</sup>

Yashar Hezaveh<sup>1,2</sup>

Laurence Perreault Levasseur<sup>1,2,3</sup>

Benjamin Wandelt<sup>2,4</sup>

<sup>1</sup>Université de Montréal

<sup>2</sup>Flatiron Institute

<sup>3</sup>Mila

<sup>4</sup>Sorbonne Université

\*ronan.legin@umontreal.ca

## Abstract

In the coming years, a new generation of sky surveys, in particular, Euclid Space Telescope (2022), and the Rubin Observatory's Legacy Survey of Space and Time (LSST, 2023) will discover more than 200,000 new strong gravitational lenses, which represents an increase of more than two orders of magnitude compared to currently known sample sizes [1]. Accurate and fast analysis of such large volumes of data under a statistical framework is therefore crucial for all sciences enabled by strong lensing. Here, we report on the application of simulation-based inference methods, in particular, density estimation techniques, to the predictions of the set of parameters of strong lensing systems from neural networks. This allows us to explicitly impose desired priors on lensing parameters, while guaranteeing convergence to the optimal posterior in the limit of perfect performance.

## 1 Introduction

Strong gravitational lensing is a phenomenon in which the light rays of distant galaxies are deflected by the gravity of foreground matter, resulting in the production of multiple images. It is a powerful probe that can map the inner distribution of matter in individual lens galaxies to reveal invaluable information about the physics of the dark matter particle [2]. It can also provide precise estimates of the expansion rate of the Universe (the Hubble constant), an important measurement, given what is referred to as a current "crisis in cosmology" (a significant tension between the measurements of this parameter by multiple probes) [3]. Strong lensing can also be used to study the magnified images of background galaxies, which are typically some of the most distant galaxies of the Universe, effectively working as a natural telescope [4].

In recent years, convolutional neural networks have been shown capable of providing accurate point estimates of the parameters describing these strong lensing systems more than 10 million times faster than traditional methods [5]. Subsequent works have expanded this result to also obtain uncertainty estimates for the predictions made by neural networks. This has been done using approximate Bayesian neural networks trained with variational inference [6]. Despite their success at producing accurate measurements in controlled experiments, these procedures involve many levels of approximations (e.g., the choice of the variational distributions for the outputs and the network weights), which cannot be easily quantified or controlled, possibly resulting in biased estimates. Additionally, despite their name, they do not offer a truly Bayesian inference framework for the lens

parameters: there is no clear way to impose explicit priors on them. The fact that rigorous uncertainty quantification is crucial for the science goals enabled by strong lensing suggests that more attention should be given to the exact form in which this inference problem is formulated.

This is fortunately possible, thanks to advances in simulation-based inference methods [7, 8, 9]. In this work, by computing the mean of the approximate posteriors provided by Bayesian neural networks and treating them as compressed statistics, we show that it is possible to learn the likelihood function of the compressed statistics given the true lensing parameters from simulations. We calculate the coverage probabilities of sampled posteriors using this likelihood and show that they are accurate while not requiring any tunable calibration process of hyperparameters. This procedure is computationally efficient, allowing the posteriors of hundreds of thousands of lenses to be calculated with minimal computational resources.

## 2 Methods

**Simulations** Strong lensing simulations use the lens equation to relate the coordinates of the image plane to the coordinates of the source plane. This relation depends on the scaled deflection angle through

$$\beta = \alpha - \alpha_E; \quad (1)$$

where  $\alpha$  is computed from the mass distribution in the lensing structure.

Our simulations are made of a Singular Isothermal Ellipsoid (SIE) with added external shear for the lensing structure and an elliptical Sérsic background source. In total, seven parameters are needed to describe the lens and six for the background source. We generate simulations by randomly sampling the lens and source parameters from a uniform distribution, spanning a wide range of lensing configurations. Each simulation contains pixel-wise independent Gaussian noise with standard deviation chosen from a uniform distribution between 1% and 10% of the peak lens surface brightness. Additional information on the method used for generating strong lensing simulations can be found in [5]. Examples of our lensing simulations are shown in Fig. 3.

**Data compression** We train an approximate Bayesian neural network to predict the distribution of the lens and background source parameters from strong lensing images. Approximate BNNs can be used with variational inference to represent the marginalized posterior distribution  $p(\theta|x)$  as

$$p(\theta|x) = \int p(\theta|x; w)q(w)dw; \quad (2)$$

where  $q(w)$  is the variational distribution of the network weights. Similar to [6],  $q(w)$  is chosen to be a Bernoulli random variable multiplied by the network weights, effectively setting a random number of weights to zero. In practice, this is achieved using dropout [10] on the output of each network layer except the final layer. The distribution over the target lensing parameters  $p(\theta|x; w)$  is chosen to be a Gaussian mixture model, expressed as

$$p(\theta|x; w) = \sum_k^K \pi_k(x; w)N(\theta; \hat{\mu}_k(x; w); \hat{\Sigma}_k(x; w)); \quad (3)$$

where  $\pi_k$ ,  $\hat{\mu}_k$  and  $\hat{\Sigma}_k$  represent the weight, mean and covariance matrix for each of the  $K$  mixture components.

Sampling from the approximate marginalized posterior  $p(\theta|x)$  is done by first feeding many times the input  $x$  to the BNN in order to get different predictions for  $p(\theta|x; w)$ , where each prediction is made from a different set of weights  $w$  due to random dropout. Then, data points are drawn from these predicted distributions, forming a set of samples that cover the distribution  $p(\theta|x)$ .

We train a two-component BNN with 20% dropout rate and another with no dropout, essentially predicting a distribution over  $p(\theta|x; w)$  parametrized by a fixed set of network weights. After training, the BNNs are used to get compressed statistics  $\hat{\mu}(x)$  of lensing simulations (by computing the mean of predicted posteriors), which, along with the true simulation parameters  $\theta$ , serve as training data for the mixture density network. The main reason for obtaining our point estimates from these networks

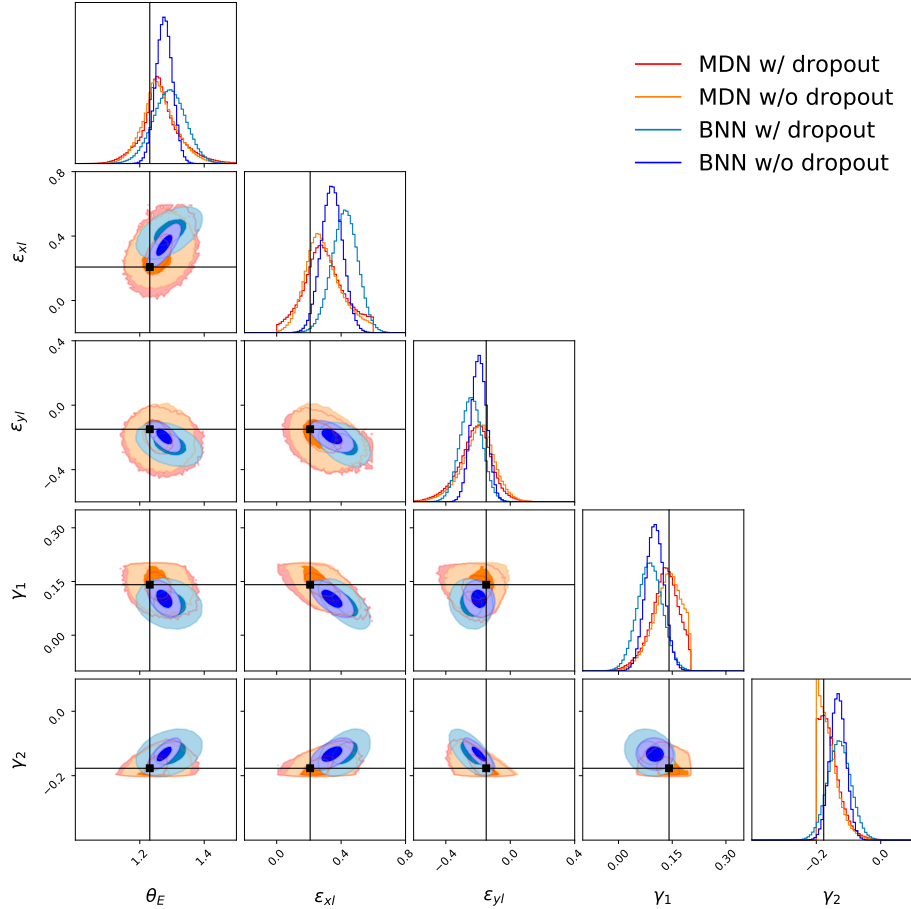


Figure 1: An example of the sampled posterior from a BNN with and without dropout, and from likelihood MDNs trained using compressed statistics of these respective BNNs. Sampling using the MDN likelihood allows for explicitly defining a prior over lens parameters, which in our case is a uniform distribution.

(instead of training a simple CNN trained with a Mean Square Error loss, for example) is to be able to compare the results of our likelihood-free inference method to approximate BNN methods for the same network, same training, and same performance.

**Density estimation** Mixture density networks (MDN) are neural networks that model conditional probability densities as a mixture of parametric distributions. Typically, the parametric model is chosen to be a mixture of Gaussian distributions similar to Eq. 3. However, an MDN models a conditional probability density parametrized by a fixed set of trained network weights  $\mathcal{W}$ , instead of assuming a variational distribution  $q(\mathcal{W})$  over the weights.

The network uses as input  $\theta$  and the Gaussian noise standard deviation used in the simulations and outputs a prediction for the parameters  $\gamma_1$ ,  $\gamma_2$ , and  $\sigma$  to model the distribution over  $\hat{\gamma}$ . Consequently, these parameters are conditioned on the input given to the network, which, in this case, gives us a model for the likelihood. The network architecture is a simple two layer feedforward neural network predicting a two-component mixture model. In order to sample from the posterior, we use this modelled likelihood with the affine-invariant MCMC algorithm implemented in the Python package PYDELFI [11]. The algorithm is written entirely in Tensorflow, allowing for GPU accelerated MCMC sampling.

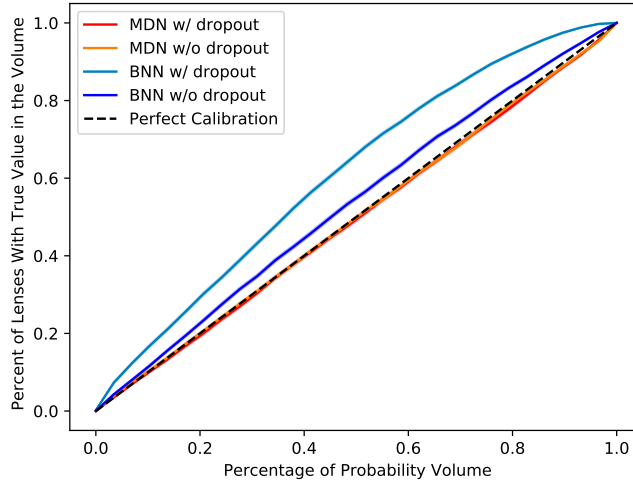


Figure 2: Results from the accuracy test conducted from predicted posteriors over a set of 10000 lensing simulations. For both BNNs, with and without dropout, the predicted posteriors are considered under confident. In the case of the MDN likelihood, this test indicates that the sampled posteriors are on average perfectly calibrated.

**Accuracy test** To evaluate the accuracy of the predicted posteriors, we calculate the coverage probabilities following [6, 12]. Simply put, the test is based on the idea that for an accurate posterior, the true value should be found  $x\%$  of the time in a region that encompasses  $x\%$  of the total probability.

To perform this test, we generate a set of 10000 lensing simulations and get samples from their predicted posterior distributions. For each simulation, we compute the distance between the true and the predicted lens parameters  $\hat{x}$  based on a chosen distance metric. Similar to [12], we define this distance metric as

$$d(x) = (x - \hat{x})^T \Sigma_d^{-1} (x - \hat{x}); \quad (4)$$

where  $\Sigma_d$  is the empirical covariance matrix of true lensing parameters over our entire set of simulations. For every simulation, we compute the fraction of posterior samples that fall within the distance between the true and mean predicted lens parameters. This gives us an estimate of the probability volume needed to include the true value.

We then check over the entire set of simulations the number of times the truth is found within some arbitrary probability volume  $x\%$ . If the truth falls within this volume  $x\%$  of the time, then the predicted posteriors are on average perfectly calibrated. Results of this test are shown in 2.

### 3 Results and discussion

An example of predicted posterior distributions of lens parameters marginalized over the source parameters is shown in Fig. 1. Although both BNN posteriors appear to be biased, the method using the modelled likelihood remains accurate, despite being trained on compressed statistics from these BNNs. This makes density estimation a powerful alternative for inference: the model learns a direct mapping between  $x$  and  $\hat{x}$ . Therefore, even if the compressed statistics are heavily biased due to the BNN having inadequate properties (e.g., using a simplistic architecture, badly calibrated dropout rate or under training the network), accurate posteriors can be obtained given that the true distribution is well modelled by the chosen parametric distribution. Fig. 2 shows the results of coverage probability tests, indicating near perfect accuracy for the simulation-based method without the need for any manual tuning.

Although sampling the posterior using an MDN likelihood required running an MCMC, this proved to be extremely fast. On average, we are capable of running MCMC chains for 200 different posterior predictions in parallel, each sampling approximately  $10^4$  samples per second using a single NVIDIA

V100 GPU. Having  $10^5$  posterior samples for each of the 10000 lensing simulations used in Fig. 1 took approximately 20 minutes. This shows that LFI density estimation of predicted lensing parameters is well suited for inferring the posterior of tens of thousands of upcoming lenses in a quick and accurate manner.

Although the SIE model has been extremely successful in representing real lensing galaxies [13], background sources can have much more complicated morphologies than the elliptical Sérsic model. This means that the number of parameters needed to describe the source may be too large for MDNs. In future work, we aim to address this issue by training a generative model for the reconstruction of background sources, leveraging the ability to sample many realizations of the source, which we can use to marginalize over it when modelling the distribution of the SIE lens parameters.

It is important to note that the simulations used in this paper did not include foreground effects often present in real strong lensing data such as stellar light from the main lens. Traditionally, these have been often masked out when sampling an explicit closed-form for the posteriors, potentially resulting in biased estimates. In simulation-based inference, these artifacts would simply be added to the simulations used to train the BNN in order to include their contributions in the final predicted posterior. Although the inference of the lensing compressed statistics may be less precise, the method used here should still converge towards accurate posteriors.

In the future, we will further evaluate the performance of our model beyond the accuracy test shown in Fig. 2. For this, we plan on using the multidimensional Kolmogorov-Smirnov (KS) test proposed in [14]. Generally, the KS test in dimensions higher than one is ill-defined. The method developed in [14] leverages the fact that the total probability mass contained in a region of highest probability density (HPD) is a unique one-dimensional statistic. Any HPD region is defined by the boundary where the probability values within are greater than at some sample  $x$ . The key idea is that the values of  $f(x)$  for repeated samples  $x$  follow a uniform distribution if  $x$  is sampled from the probability distribution where the HPD region is taken from. For our purpose, we can use this to verify whether the true lensing parameters can be accurately sampled from the predicted posteriors. This is done by means of a simple one-dimensional KS test which verifies the null hypothesis that computed using our predicted posteriors follows a uniform distribution.

We would like to emphasize that the purpose of obtaining compressed statistics from a BNN rather than a deterministic CNN is so that we may compare its predicted posteriors with the ones obtained using the density estimation method. However, it is unclear at the moment if using compressed statistics based on samples obtained from BNNs with a non-zero dropout rate could make the density estimation task more difficult: it may be necessary to consider a larger MDN training set or a more expressive likelihood model to better capture the potential added noise that dropout might induce in the compressed statistics. For now, this does not seem to be the case based on Fig. 2: the two MDN models trained on compressed statistics from 0% and 20% dropout BNNs showed on average similar levels of accuracy. The effects of dropout on the MDN-based posteriors will be further explored as we plan on evaluating the performance of our method on additional tests such as the multidimensional KS test described above.

## 4 Broader Impact

This work is contributing to the development of uncertainty estimation methods for the accurate analysis of astronomical data using deep learning. The proposed approach could be applied to other problems in other domains where accurate and detailed quantification of uncertainties is crucial, and its impact could be most important in cases where high-dimensional posteriors need to be characterized.

## Acknowledgments and Disclosure of Funding

This research was supported by a generous grant from the Schmidt Futures Foundation. The work was also enabled in part by computational resources provided by Calcul Quebec and Compute Canada. Y.H. and L.P. acknowledge support from the National Science and Engineering Council of Canada, the Fonds de recherche du Québec, and the Canada Research Chairs Program. We thank Adam Coogan, postdoctoral researcher at the University of Montreal, for useful discussions in regards to simulation-based inference methods.

## References

- [1] Thomas E. Collett. The Population of Galaxy-Galaxy Strong Lenses in Forthcoming Optical Imaging Surveys. , 811(1):20, September 2015.
- [2] Yashar D. Hezaveh, Neal Dalal, Daniel P. Marrone, Yao-Yuan Mao, Warren Morningstar, Di Wen, Roger D. Blandford, John E. Carlstrom, Christopher D. Fassnacht, Gilbert P. Holder, Athol Kembball, Philip J. Marshall, Norman Murray, Laurence Perreault Levasseur, Joaquin D. Vieira, and Risa H. Wechsler. Detection of Lensing Substructure Using ALMA Observations of the Dusty Galaxy SDP.81. , 823(1):37, May 2016.
- [3] Kenneth C. Wong, Sherry H. Suyu, Geoff C. F. Chen, Cristian E. Rusu, Martin Millon, Dominique Sluse, Vivien Bonvin, Christopher D. Fassnacht, Stefan Taubenberger, Matthew W. Auger, Simon Birrer, James H. H. Chan, Frederic Courbin, Stefan Hilbert, Olga Tihhonova, Tommaso Treu, Adriano Agnello, Xuheng Ding, Inh Jee, Eiichiro Komatsu, Anowar J. Shajib, Alessandro Sonnenfeld, Roger D. Blandford, Léon V. E. Koopmans, Philip J. Marshall, and Georges Meylan. H0LiCOW – XIII. A 2.4 per cent measurement of  $H_0$  from lensed quasars: 5.3 tension between early- and late-Universe probes. , 498(1):1420–1439, October 2020.
- [4] Dan Coe, Adi Zitrin, Mauricio Carrasco, Xinwen Shu, Wei Zheng, Marc Postman, Larry Bradley, Anton Koekemoer, Rychard Bouwens, Tom Broadhurst, Anna Monna, Ole Host, Leonidas A. Moustakas, Holland Ford, John Moustakas, Arjen van der Wel, Megan Donahue, Steven A. Rodney, Narciso Benítez, Stephanie Jouvel, Stella Seitz, Daniel D. Kelson, and Piero Rosati. CLASH: Three Strongly Lensed Images of a Candidate  $z \approx 11$  Galaxy. , 762(1):32, January 2013.
- [5] Yashar D. Hezaveh, Laurence Perreault Levasseur, and Philip J. Marshall. Fast automated analysis of strong gravitational lenses with convolutional neural networks. , 548(7669):555–557, August 2017.
- [6] Laurence Perreault Levasseur, Yashar D. Hezaveh, and Risa H. Wechsler. Uncertainties in Parameters Estimated with Neural Networks: Application to Strong Gravitational Lensing. , 850(1):L7, November 2017.
- [7] Kyle Cranmer, Johann Brehmer, and Gilles Louppe. The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, 117(48):30055–30062, 2020.
- [8] Johann Brehmer, Siddharth Mishra-Sharma, Joeri Hermans, Gilles Louppe, and Kyle Cranmer. Mining for Dark Matter Substructure: Inferring Subhalo Population Properties from Strong Lenses with Machine Learning. , 886(1):49, November 2019.
- [9] Adam Coogan, Konstantin Karchev, and Christoph Weniger. Targeted Likelihood-Free Inference of Dark Matter Substructure in Strongly-Lensed Galaxies. *arXiv e-prints*, page arXiv:2010.07032, October 2020.
- [10] Yarin Gal and Zoubin Ghahramani. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. *arXiv e-prints*, page arXiv:1506.02142, June 2015.
- [11] Justin Alsing, Tom Charnock, Stephen Feeney, and Benjamin Wandelt. Fast likelihood-free cosmology with neural density estimators and active learning. , 488(3):4440–4458, September 2019.
- [12] Sebastian Wagner-Carena, Ji Won Park, Simon Birrer, Philip J. Marshall, Aaron Roodman, Risa H. Wechsler, and LSST Dark Energy Science Collaboration. Hierarchical Inference with Bayesian Neural Networks: An Application to Strong Gravitational Lensing. , 909(2):187, March 2021.
- [13] Adam S. Bolton, Scott Burles, Leon V. E. Koopmans, Tommaso Treu, Raphael Gavazzi, Leonidas A. Moustakas, Randall Wayth, and David J. Schlegel. The Sloan Lens ACS Survey. V. The Full ACS Strong-Lens Sample. , 682(2):964–984, August 2008.
- [14] Diana Harrison, David Sutton, Pedro Carvalho, and Michael Hobson. Validation of Bayesian posterior distributions using a multidimensional Kolmogorov-Smirnov test. , 451(3):2610–2624, August 2015.

## Checklist

1. For all authors...
  - (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
  - (b) Did you describe the limitations of your work? [Yes] See final paragraph in section 3 "Results and discussion".
  - (c) Did you discuss any potential negative societal impacts of your work? [No]
  - (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
  - (a) Did you state the full set of assumptions of all theoretical results? [N/A]
  - (b) Did you include complete proofs of all theoretical results? [N/A]
3. If you ran experiments...
  - (a) 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]
  - (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [No] Additional details regarding the training of Bayesian neural networks and on the generation of lensing simulations can be found in [6] and [5].
  - (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [No]
  - (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [No] We only reported the amount of computation time when running an MCMC using the MDN likelihood in section 3. However, throughout our work we used a single NVIDIA V100 GPU.
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
  - (a) If your work uses existing assets, did you cite the creators? [Yes] See section 2 under "Density estimation".
  - (b) Did you mention the license of the assets? [No]
  - (c) Did you include any new assets either in the supplemental material or as a URL? [No]
  - (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [No]
  - (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [No]
5. If you used crowdsourcing or conducted research with human subjects...
  - (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
  - (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
  - (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

## A Appendix

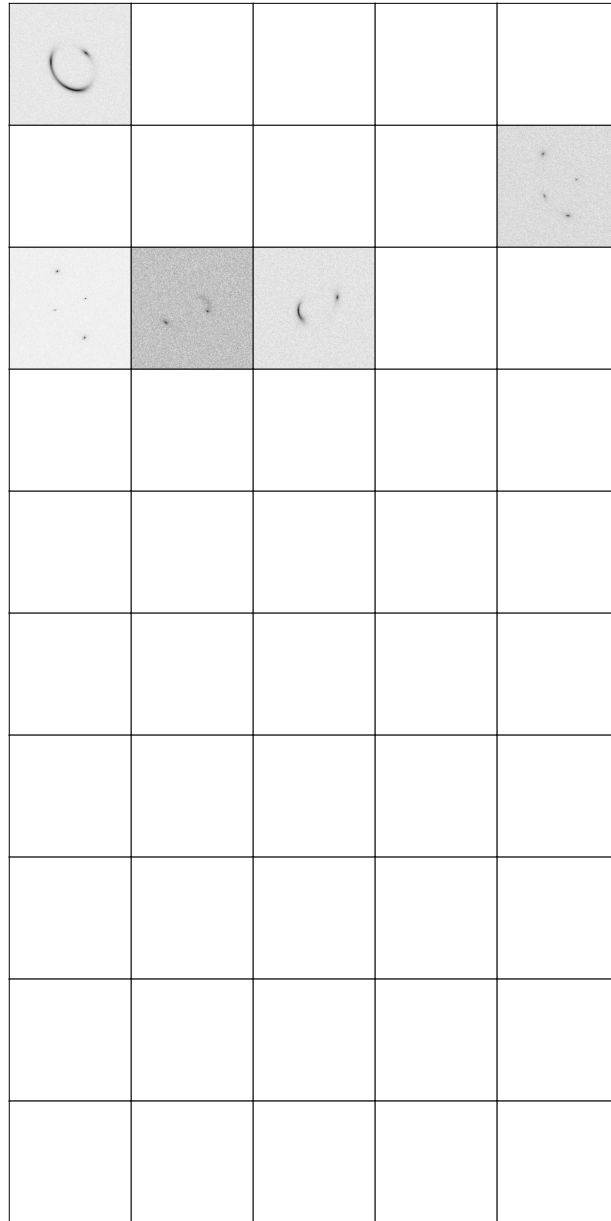


Figure 3: Examples of generated lensing simulations.