
Population-Level Inference of Strong Gravitational Lenses with Neural Network-Based Selection Correction

Ronan Legin¹ Connor Stone^{1 2 3} Yashar Hezaveh^{1 4} Laurence Perreault-Levasseur^{1 3 4}

Abstract

A new generation of sky surveys is poised to provide unprecedented volumes of data containing hundreds of thousands of new strong lensing systems in the coming years. Convolutional neural networks are currently the only state-of-the-art method that can handle the onslaught of data to discover and infer the parameters of individual systems. However, many important measurements that involve strong lensing require population-level inference of these systems. In this work, we propose a hierarchical inference framework that uses the inference of individual lensing systems in combination with the selection function to estimate population-level parameters. In particular, we show that it is possible to model the selection function of a CNN-based lens finder with a neural network classifier, enabling fast inference of population-level parameters without the need for expensive Monte Carlo simulations.

1. Introduction

Strong gravitational lensing is a phenomenon where light rays of a distant background source are deflected by the gravity of foreground matter, resulting in the production of multiple images. These lensing systems can be used for important applications in cosmology, such as constraining the mass of dark matter present in foreground galaxies (e.g., [Vegetti et al., 2012](#); [Hezaveh et al., 2016](#)), providing precise measurements of the expansion rate of the universe (e.g., [Wong et al., 2020](#)) and learning more about the morphology of distant high redshift galaxies (e.g., [Coe et al., 2013](#)).

¹Department of Physics, Université de Montréal, Montréal, Québec, Canada ²Department of Physics, Engineering Physics and Astronomy, Queen's University, Kingston, ON K7L 3N6, Canada ³Mila - Quebec Artificial Intelligence Institute, Montréal, Québec, Canada ⁴Center for Computational Astrophysics, Flatiron Institute, NY, USA. Correspondence to: Ronan Legin <ronan.legin@umontreal.ca>.

In the next few years, a new generation of wide-area sky surveys from the Rubin Observatory's Legacy Survey of Space and Time (LSST) and the Euclid Space Telescope (EST) will provide an unprecedented volume of data containing hundreds of thousands of new strong lensing systems ([Collett, 2015](#)).

Recent works have focused on developing new and efficient computational methods to analyze these large sets of data. In particular, machine learning has been proposed to efficiently perform various strong lensing analysis tasks. Convolutional neural networks (CNN) have been trained in classification tasks to detect strong lenses from survey data ([Schaefer et al., 2018](#); [Lanusse et al., 2018](#)). They have been used to discover new strong lensing systems from the Kilo Degree Survey ([Li et al., 2020](#), KiDS) and the Ultraviolet Near Infrared Optical Northern Survey ([Savary et al., 2021](#)), and have generally surpassed traditional methods in terms of speed and accuracy according to benchmark tests ([Metcalf et al., 2019](#)). These networks have also been used to perform fast and automated inference of the lensing parameters from the observations of individual strong lensing systems ([Hezaveh et al., 2017](#); [Perreault Levasseur et al., 2017](#); [Wagner-Carena et al., 2021](#); [Park et al., 2021](#); [Legin et al., 2021](#)), speeding up the analysis by more than ~ 10 million times compared to traditional methods.

However, many important science applications involving strong gravitational lenses rely on population-level statistics of lensing systems (e.g., [Sonnenfeld & Cautun, 2021](#); [Sonnenfeld, 2021; 2022b;a](#)). Generally, the task of population-level inference is best framed within a hierarchical framework, where the problem is broken down in multiple stages. In a hierarchical framework, a clear separation between the different components of the analysis is provided, (e.g., between the theoretical models underlying the population distribution and the parameters describing the individual observations themselves). Generally, the problem includes the inference of global hyperparameters λ describing the population distribution of model parameters θ i.e the parameters describing individual strong lenses. It is important, however, for these hierarchical models to take into account the selection mechanisms of the observed data sets to avoid biased inference results ([Sonnenfeld & Cautun, 2021](#); [Sonnenfeld, 2022a](#)). The selection function would depend on the model

parameters of individual strong lensing observations. Traditionally, this makes population-level inference difficult in standard settings, where complex selections are difficult to quantify and the corrections require a large number of simulations at every step of the posterior sampling process.

In this work, we model the selection function of a CNN-based strong lensing detector using neural networks. We propose an efficient and flexible likelihood-free inference framework for performing accurate large scale population-level inference of strong lensing parameters that can readily incorporate learned selection effects. In addition, our framework circumvents issues regarding the interpretability of typical black-box inference machines by allowing clear separation between the different components of our hierarchical population model. In section 2, we describe our Bayesian hierarchical inference framework for inferring the posterior distribution of hyperparameters λ describing the population of strong gravitational lensing model parameters θ . We also detail our strong lensing simulations and the modeling of the CNN-based selection bias using neural networks. In section 3, we test our hierarchical method to infer the population of strong lensing model parameters based on a set of 1000 simulated strong lensing observations.

2. Methods

Given a set of observations $\{x_i\}$, the task is to infer the posterior distribution of population hyperparameters $p(\lambda|\{x_i\})$. Using Bayes' theorem, this can be expressed as

$$\begin{aligned} p(\lambda|\{x_i\}) &= \frac{p(\lambda) \prod_i p(x_i|\lambda)}{\int d\lambda' p(\lambda') \prod_i p(x_i|\lambda')} \\ &= \left[\int d\lambda' \frac{p(\lambda')}{p(\lambda)} \prod_i \frac{p(x_i|\lambda')}{p(x_i|\lambda)} \right]^{-1}, \end{aligned} \quad (1)$$

In a typical hierarchical model, the individual observations x_i may explicitly depend on intermediate model parameters θ and only implicitly on λ . In this case, the likelihood $p(x_i|\lambda)$ can be expanded as

$$p(x_i|\lambda) = \int p(x_i|\theta)p(\theta|\lambda)d\theta, \quad (2)$$

where λ is ignored in $p(x_i|\theta)$ as it only implicitly affects the observation x_i .

In practice, it may be more likely to observe certain events x_i . This may lead to the chosen set of events $\{x_i\}$ to instead represent a biased selection of all possible observations, which would consequently result in a biased inference of the posterior distribution of the population hyperparameters

$p(\lambda|\{x_i\})$. In the presence of selection effects (Mandel et al., 2019; Moore & Gerosa, 2021), the proper expression for the likelihood $p(x_i|\lambda)$ is

$$p(x_i|\lambda) = \frac{\int p(x_i|\theta)p(\theta|\lambda)d\theta}{\alpha(\lambda)}, \quad (3)$$

where $\alpha(\lambda)$ is evaluated by integrating over the product of the probability of detection $p_{\text{det}}(\theta)$ of an event with model parameters θ and the population model $p(\theta|\lambda)$,

$$\alpha(\lambda) = \int p_{\text{det}}(\theta)p(\theta|\lambda)d\theta. \quad (4)$$

As such, accounting for selection bias incorporates an overall normalization factor $\alpha(\lambda)$ within the likelihood $p(x|\lambda)$ that can be interpreted as the fraction of samples from the population model $p(\theta|\lambda)$ that would be observed given the selection function $p_{\text{det}}(\theta)$ (Mandel et al., 2019). Note that Equation 3 can be rewritten as

$$p(x_i|\lambda) = \mathbb{E}_{p(\theta|x_i)} \left[\frac{p(\theta|\lambda)}{p(\theta)} \right] \frac{p(x_i)}{\alpha(\lambda)}, \quad (5)$$

by applying Bayes' theorem on the likelihood of the individual lensing observations $p(x_i|\theta)$ to obtain an expectation value over the lensing model parameter posterior $p(\theta|x_i)$.

Plugging this new expression back into Equation 1, we obtain

$$p(\lambda|\{x_i\}) = \left[\int d\lambda' \frac{p(\lambda')}{p(\lambda)} \prod_i \frac{\mathbb{E}_{p(\theta|x_i)} \left[\frac{p(\theta|\lambda')}{p(\theta)\alpha(\lambda')} \right]}{\mathbb{E}_{p(\theta|x_i)} \left[\frac{p(\theta|\lambda)}{p(\theta)\alpha(\lambda)} \right]} \right]^{-1}, \quad (6)$$

where the marginal probabilities $p(x_i)$ from applying Bayes' theorem on both $p(x_i|\lambda)$ and $p(x_i|\lambda')$ cancel out. As the expectation values cannot be evaluated analytically, we can instead approximate them as a mean over samples from the lensing posterior $p(\theta|x_i)$. Similarly, we can estimate $\alpha(\lambda)$ by rewriting Equation 4 as the expectation value of $p_{\text{det}}(\theta)$ over $p(\theta|\lambda)$ and approximate it with samples from $p(\theta|\lambda)$.

The difficulty in estimating $\alpha(\lambda)$, however, is that the selection function $p_{\text{det}}(\theta)$, which captures the effects of the selection bias, is typically complex and not explicitly known (e.g., Sonnenfeld, 2022a). This is especially true if the selection method is based on CNN detections of strong lenses, as the entire detection algorithm is framed as a black-box.

In this work, we circumvent this issue by modeling the selection function $p_{\text{det}}(\theta)$ of CNN-based detectors of strong

lenses. In particular, we train a neural network to classify between detected and non-detected samples of strong lensing model parameters θ . The output probability of this classification network results in the probability of detection $p_{\text{det}}(\theta)$ of a lensing system with parameters θ .

Furthermore, we train a convolutional Bayesian Neural Network (BNN) to predict the posterior distribution of strong lensing model parameters $p(\theta|x)$ and to rapidly sample from it in order to approximate the expectation values from Equation 6. Given that upcoming sky surveys are estimated to discover hundreds of thousands of new strong lenses, traditional methods such as likelihood-based inference, which require time-consuming analyses from strong lensing experts, will be intractable for this task. Instead, BNNs have been shown to be orders of magnitude faster than traditional methods and can provide well-calibrated uncertainties on strong lensing parameter predictions (Perreault Levasseur et al., 2017; Wagner-Carena et al., 2021; Park et al., 2021).

2.1. Simulations

We simulate two sets of strong lensing observations: low-resolution images similar to what is expected to be observed by wide-area surveys such as LSST and the corresponding high-resolution follow-up observations. As a whole, our simulations can be described by a 13-dimensional model parameter vector θ . Examples of the lensing simulations are shown in appendix A.

At test time, the CNN detector is applied to the set of mock LSST lensing observations. Then, for the images that are classified as a detection, we apply our BNN model to their follow-up observations in order to infer the posterior distribution of strong lensing model parameters $p(\theta|x)$ with higher precision. Given the inferred posterior $p(\theta|x)$, we can evaluate Equation 6 by approximating the expectation values using samples from it.

2.2. Selection function

We train a multilayer perceptron (MLP) neural network to predict the probability of detection $p_{\text{det}}(\theta)$ for CNN-based detectors of strong gravitational lenses. This neural network is trained to classify between a set of θ that results in the detection of strong lensing systems and a set that results in a non-detection.

To generate training data, we apply a CNN detector to a set of simulated low-resolution strong lensing images. This reflects the lens search procedure in LSST-like survey data. Lenses with parameters θ which are detected by the CNN are classified as detectable, and the missed lenses are classified as non-detectable. An illustration of the training data generation is shown in Figure 1.

The neural network that predicts the selection function

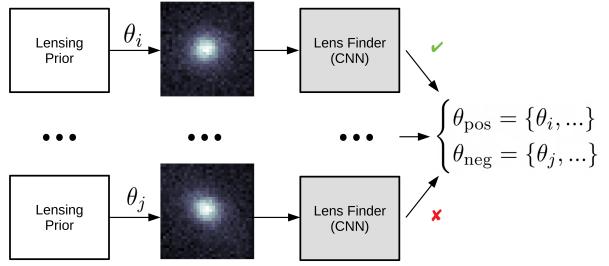


Figure 1. Flow chart illustrating the generation of training data used for training the neural selection function. The CNN-based lens finder is applied to simulated low-resolution strong lensing images with model parameters θ sampled from a prior distribution. Two sets of model parameters are made based on whether the images were detected as strong lenses. The neural selection function is trained to classify between these two sets, where the output classification probability represents the probability of detection $p_{\text{det}}(\theta)$.

$p_{\text{det}}(\theta)$ is composed of an input layer and a hidden layer each with 256 neurons, in addition to a final layer outputting a single value. We use the Exponential Linear Unit activation function (Clevert et al., 2015, ELU) for the first two layers, and the sigmoid activation for the output of the network. The final layer predicts the probability of θ being classified as detectable, which gives us a prediction for the probability of detection $p_{\text{det}}(\theta)$. The weights of the neural network are updated using the Adam optimizer with a constant learning rate of 10^{-4} . Our training data consists of two equal sets of 10,000 detectable and non-detectable θ parameters. The network is trained for 10,000 epochs with a batch size of 1000.

3. Results

We compute the posterior $p(\lambda|\{x_i\})$ using Equation 6 from a set of 1000 CNN-detected strong lensing observations. As detailed in section 2.2, we first train a neural network to predict the CNN detector selection function p_{det} . Then, we use the neural network's prediction of p_{det} in Equation 4 to compute the normalization factor $\alpha(\lambda)$ over a grid of λ values. Separately, we train a BNN to infer the lensing posterior $p(\theta|x_i)$ for the individual strong lensing observations. In the final step, we use our prediction of $\alpha(\lambda)$ and $p(\theta|x_i)$ to compute the posterior of population hyperparameters $p(\lambda|\{x_i\})$ from Equation 6. To approximate the expectation values from Equation 6, we use 100,000 samples from the posterior distribution $p(\theta|x_i)$ for each observation x_i . We compute $\alpha(\lambda)$ by approximating the integral from Equation 4 using 10,000 samples from the population distribution $p(\theta|\lambda)$. The computation of $\alpha(\lambda)$ is repeated 1000 times

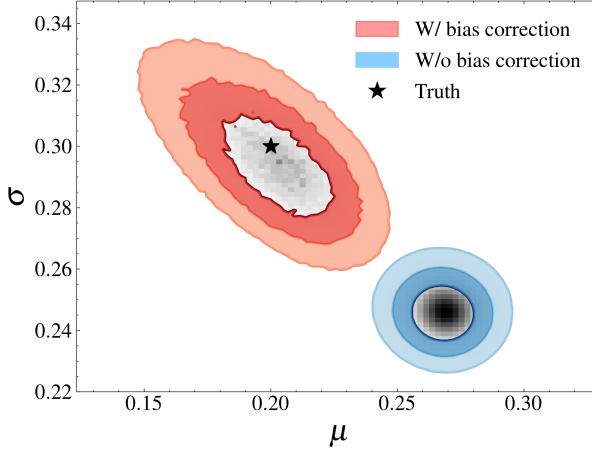


Figure 2. Predicted posterior $p(\lambda|x_i)$ given a set of 1000 strong lensing observations detected using the CNN strong lensing detector with the selection bias correction included within the hierarchical model (red contours) and without (blue contours). The 68%, 95% and 99.7% probability regions are shown for both posterior distributions. This result demonstrates that ignoring the effects of the selection bias caused by the CNN detectors leads to a biased inference of the population hyperparameters λ .

using different samples from $p(\theta|\lambda)$ and averaged over in order to reduce noise in its prediction. This helps mitigate instabilities that may occur when dividing by the normalization factor $\alpha(\lambda)$ in Equation 6. A flow chart illustrating the problem setup for inference is shown in Figure 5.

For our problem, the hyperparameters λ that we wish to infer are the mean μ and standard deviation σ of a log-normal distribution describing the population distribution of strong lensing Einstein radii R_E . As an initial test, we apply a mask on the strong lensing light for lensing systems with Einstein radius smaller than 0.8 arcseconds, making it impossible for the CNN finder to detect the presence of strong lensing. This leads to a more important selection bias towards systems with larger Einstein radii, and for the initial test, allows to more intuitively visualize the impact of selection bias on the final result. In Figure 2, we show results of the predicted posterior $p(\lambda|x_i)$ for two different scenarios: one in which the hierarchical model includes the correction due to selection bias and the other where it is ignored (i.e. whether $\alpha(\lambda)$ is included in Equation 6).

We also perform a test to verify if our modeled selection function p_{det} accurately portrays the performance of the CNN strong lensing detector. To do so, we start by generating a new set of 10,000 strong lensing observations based on a wide range of model parameters θ , including the Einstein radius R_E . Then, in the first case, we evaluate the detection probability $p_{\text{det}}(\theta)$ for these θ samples, and keep the samples considered as detected if they satisfy $\rho < p_{\text{det}}(\theta)$, where ρ is a number randomly sampled between zero and

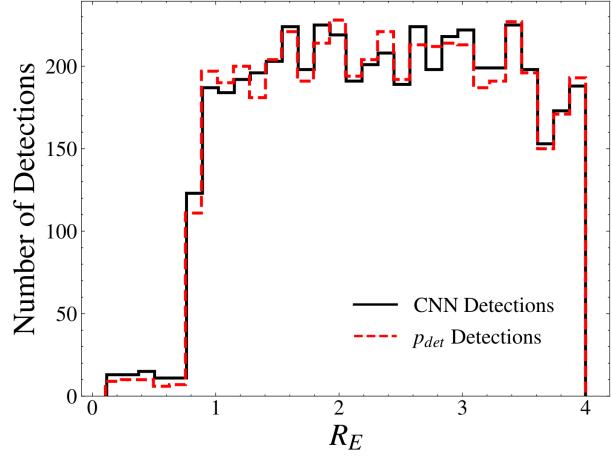


Figure 3. The predicted number of strong lenses detected from a set of 10,000 simulated strong lenses by the CNN detector (black line) and by our modeled selection function p_{det} (dotted red line) as a function of the Einstein radius R_E . The neural selection function p_{det} is capable of accurately portraying the selection mechanism of CNN-based lens detectors.

one for each θ . We denote this set of detected θ samples as $\theta_{p_{\text{det}}}$. In the second case, we apply the CNN detector directly on the strong lensing images. For the observations that passed as a detection, we keep the samples of θ that were used to generate them. We denote this set of samples as θ_{CNN} . In Figure 3, we plot the distribution of the Einstein radius parameter R_E from both sets of $\theta_{p_{\text{det}}}$ and θ_{CNN} . This test allows us to verify if our modeled selection function $p_{\text{det}}(\theta)$ accurately reflects the overall distribution of strong lenses that are discovered by the CNN detector.

4. Discussion

The framework presented in this paper allows for the hierarchical inference of population-level parameters of strong lensing systems, including the effects of selection functions. In the first stage of inference, the posteriors of the parameters of individual lensing systems are obtained. In the second stage, the distribution of the parameters of the general population of lenses is inferred using the posteriors of individual lens parameters. In this second stage of inference, it is also essential to include the effect of the selection function.

It is of course possible to forgo a hierarchical framework and infer population-level parameters in a one-stage inference setup. Simulation-based (or likelihood-free) inference methods that directly model target quantities such as $p(x|\lambda)$ have proven to be promising methods for inferring population-level parameters in cosmology (Brehmer et al., 2019; Coogan et al., 2020; Gerardi et al., 2021).

However, Bayesian hierarchical modeling can be a flexible

and powerful alternative, providing several advantages. A distinct advantage of hierarchical modeling is the fact that the inference of the population-level parameters is done independently from the inference of individual systems. In this framework, the posterior of the lens parameters given the data $p(\theta|x_i)$ (the noise term) can be calculated once with expert (lensing) knowledge, either using machine learning based methods (e.g., Hezaveh et al., 2017; Wagner-Carena et al., 2021; Legin et al., 2021) or likelihood-based methods (e.g., Gu et al., 2022). The inference of population-level parameters could then be carried out and explored without redoing any analysis of the observed data. The term $p(\theta|\lambda)$, often called the “theory term”, describes the theoretical prediction of lens parameters under a proposed hyperparameter. This allows one to explore any theoretical model in a modular way using the same set of posteriors, $p(\theta|x_i)$, calculated earlier.

In this work, we modeled the selection function $p_{\text{det}}(\theta)$ with neural networks, facilitating accurate calculation of the posterior of the hyperparameters without the need for repeated Monte Carlo simulations. Once learned, the same neural selection function could be used to recalculate the denominator of Equation 3 for any new theoretical model under consideration.

The approach proposed here uses simulation-based inference and machine learning models to learn surrogate models for the various modules that form a hierarchical inference model. This allows taking advantage of the power and the versatility of neural networks while retaining the benefits and the interpretability of Bayesian hierarchical frameworks, where uncertainties are correctly propagated from one level to the next.

Acknowledgements

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, Compute Canada and the Digital Research Alliance of Canada. Y.H. and L.P. acknowledge support from the National Sciences and Engineering Council of Canada grant RGPIN-2020-05102, the Fonds de recherche du Québec grant 2022-NC-301305, and the Canada Research Chairs Program. We acknowledge support from the Natural Sciences and Engineering Research Council of Canada (NSERC), [funding reference number 411299547]. We thank Adam Coogan and Alexandre Adam for useful discussions in regards to simulation-based inference methods.

References

Brehmer, J., Mishra-Sharma, S., Hermans, J., Louppe, G.,

and Cranmer, K. Mining for Dark Matter Substructure: Inferring Subhalo Population Properties from Strong Lenses with Machine Learning. *ApJ*, 886(1):49, November 2019. doi: 10.3847/1538-4357/ab4c41.

Clevert, D.-A., Unterthiner, T., and Hochreiter, S. Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs). *arXiv e-prints*, art. arXiv:1511.07289, November 2015.

Coe, D., Zitrin, A., Carrasco, M., Shu, X., Zheng, W., Postman, M., Bradley, L., Koekemoer, A., Bouwens, R., Broadhurst, T., Monna, A., Host, O., Moustakas, L. A., Ford, H., Moustakas, J., van der Wel, A., Donahue, M., Rodney, S. A., Benítez, N., Jouvel, S., Seitz, S., Kelson, D. D., and Rosati, P. CLASH: Three Strongly Lensed Images of a Candidate $z \approx 11$ Galaxy. *ApJ*, 762(1):32, January 2013. doi: 10.1088/0004-637X/762/1/32.

Collett, T. E. The Population of Galaxy-Galaxy Strong Lenses in Forthcoming Optical Imaging Surveys. *ApJ*, 811(1):20, September 2015. doi: 10.1088/0004-637X/811/1/20.

Coogan, A., Karchev, K., and Weniger, C. Targeted Likelihood-Free Inference of Dark Matter Substructure in Strongly-Lensed Galaxies. *arXiv e-prints*, art. arXiv:2010.07032, October 2020.

Gerardi, F., Feeney, S. M., and Alsing, J. Unbiased likelihood-free inference of the Hubble constant from light standard sirens. *Phys. Rev. D*, 104(8):083531, October 2021. doi: 10.1103/PhysRevD.104.083531.

Gu, A., Huang, X., Sheu, W., Aldering, G., Bolton, A. S., Boone, K., Dey, A., Filippenko, A., Jullo, E., Perlmutter, S., Rubin, D., Schlafly, E. F., Schlegel, D. J., Shu, Y., and Suyu, S. H. GIGA-Lens: Fast Bayesian Inference for Strong Gravitational Lens Modeling. *arXiv e-prints*, art. arXiv:2202.07663, February 2022.

Hezaveh, Y. D., Dalal, N., Marrone, D. P., Mao, Y.-Y., Morningstar, W., Wen, D., Blandford, R. D., Carlstrom, J. E., Fassnacht, C. D., Holder, G. P., Kemball, A., Marshall, P. J., Murray, N., Perreault Levasseur, L., Vieira, J. D., and Wechsler, R. H. Detection of Lensing Substructure Using ALMA Observations of the Dusty Galaxy SDP.81. *ApJ*, 823(1):37, May 2016. doi: 10.3847/0004-637X/823/1/37.

Hezaveh, Y. D., Perreault Levasseur, L., and Marshall, P. J. Fast automated analysis of strong gravitational lenses with convolutional neural networks. *Nature*, 548(7669): 555–557, August 2017. doi: 10.1038/nature23463.

Kormann, R., Schneider, P., and Bartelmann, M. Isothermal elliptical gravitational lens models. *A&A*, 284:285–299, April 1994.

- Lanusse, F., Ma, Q., Li, N., Collett, T. E., Li, C.-L., Ravanbakhsh, S., Mandelbaum, R., and Póczos, B. CMU DeepLens: deep learning for automatic image-based galaxy-galaxy strong lens finding. *MNRAS*, 473(3):3895–3906, January 2018. doi: 10.1093/mnras/stx1665.
- Legin, R., Hezaveh, Y., Perreault Levasseur, L., and Wandelt, B. Simulation-Based Inference of Strong Gravitational Lensing Parameters. *arXiv e-prints*, art. arXiv:2112.05278, December 2021.
- Li, R., Napolitano, N. R., Tortora, C., Spiniello, C., Koopmans, L. V. E., Huang, Z., Roy, N., Vernardos, G., Chatterjee, S., Giblin, B., Getman, F., Radovich, M., Covone, G., and Kuijken, K. New High-quality Strong Lens Candidates with Deep Learning in the Kilo-Degree Survey. *ApJ*, 899(1):30, August 2020. doi: 10.3847/1538-4357/ab9dfa.
- Mandel, I., Farr, W. M., and Gair, J. R. Extracting distribution parameters from multiple uncertain observations with selection biases. *MNRAS*, 486(1):1086–1093, June 2019. doi: 10.1093/mnras/stz896.
- Metcalf, R. B., Meneghetti, M., Avestruz, C., Bellagamba, F., Bom, C. R., Bertin, E., Cabanac, R., Courbin, F., Davies, A., Decencière, E., Flamaray, R., Gavazzi, R., Geiger, M., Hartley, P., Huertas-Company, M., Jackson, N., Jacobs, C., Jullo, E., Kneib, J. P., Koopmans, L. V. E., Lanusse, F., Li, C. L., Ma, Q., Makler, M., Li, N., Lightman, M., Petrillo, C. E., Serjeant, S., Schäfer, C., Sonnenfeld, A., Tagore, A., Tortora, C., Tuccillo, D., Valentín, M. B., Velasco-Forero, S., Verdoes Kleijn, G. A., and Vernardos, G. The strong gravitational lens finding challenge. *A&A*, 625:A119, May 2019. doi: 10.1051/0004-6361/201832797.
- Moore, C. J. and Gerosa, D. Population-informed priors in gravitational-wave astronomy. *Phys. Rev. D*, 104(8):083008, October 2021. doi: 10.1103/PhysRevD.104.083008.
- Park, J. W., Wagner-Carena, S., Birrer, S., Marshall, P. J., Lin, J. Y.-Y., Roodman, A., and LSST Dark Energy Science Collaboration. Large-scale Gravitational Lens Modeling with Bayesian Neural Networks for Accurate and Precise Inference of the Hubble Constant. *ApJ*, 910(1):39, March 2021. doi: 10.3847/1538-4357/abdfc4.
- Perreault Levasseur, L., Hezaveh, Y. D., and Wechsler, R. H. Uncertainties in Parameters Estimated with Neural Networks: Application to Strong Gravitational Lensing. *ApJ*, 850(1):L7, November 2017. doi: 10.3847/2041-8213/aa9704.
- Savary, E., Rojas, K., Maus, M., Clément, B., Courbin, F., Gavazzi, R., Chan, J. H. H., Lemon, C., Vernardos, G., Cañameras, R., Schuldt, S., Suyu, S. H., Cuillandre, J. C., Fabbro, S., Gwyn, S., Hudson, M. J., Kilbinger, M., Scott, D., and Stone, C. A search for galaxy-scale strong gravitational lenses in the Ultraviolet Near Infrared Optical Northern Survey (UNIONS). *arXiv e-prints*, art. arXiv:2110.11972, October 2021.
- Schaefer, C., Geiger, M., Kuntzer, T., and Kneib, J. P. Deep convolutional neural networks as strong gravitational lens detectors. *A&A*, 611:A2, March 2018. doi: 10.1051/0004-6361/201731201.
- Sérsic, J. L. 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.
- Sonnenfeld, A. Statistical strong lensing. II. Cosmology and galaxy structure with time-delay lenses. *A&A*, 656:A153, December 2021. doi: 10.1051/0004-6361/202142062.
- Sonnenfeld, A. Statistical strong lensing. III. Inferences with complete samples of lenses. *A&A*, 659:A132, March 2022a. doi: 10.1051/0004-6361/202142301.
- Sonnenfeld, A. Statistical strong lensing. IV. Inferences with no individual source redshifts. *A&A*, 659:A133, March 2022b. doi: 10.1051/0004-6361/202142467.
- Sonnenfeld, A. and Cautun, M. Statistical strong lensing. I. Constraints on the inner structure of galaxies from samples of a thousand lenses. *A&A*, 651:A18, July 2021. doi: 10.1051/0004-6361/202140549.
- Vegetti, S., Lagattuta, D. J., McKean, J. P., Auger, M. W., Fassnacht, C. D., and Koopmans, L. V. E. Gravitational detection of a low-mass dark satellite galaxy at cosmological distance. *Nature*, 481(7381):341–343, January 2012. doi: 10.1038/nature10669.
- Wagner-Carena, S., Park, J. W., Birrer, S., Marshall, P. J., Roodman, A., Wechsler, R. H., and LSST Dark Energy Science Collaboration. Hierarchical Inference with Bayesian Neural Networks: An Application to Strong Gravitational Lensing. *ApJ*, 909(2):187, March 2021. doi: 10.3847/1538-4357/abdf59.
- Wong, K. C., Suyu, S. H., Chen, G. C. F., Rusu, C. E., Millon, M., Sluse, D., Bonvin, V., Fassnacht, C. D., Taubenberger, S., Auger, M. W., Birrer, S., Chan, J. H. H., Courbin, F., Hilbert, S., Tihhonova, O., Treu, T., Agnello, A., Ding, X., Jee, I., Komatsu, E., Shajib, A. J., Sonnenfeld, A., Blandford, R. D., Koopmans, L. V. E., Marshall, P. J., and Meylan, G. H0LiCOW – XIII. A 2.4 per cent measurement of H_0 from lensed quasars: 5.3σ tension between early- and late-Universe probes. *MNRAS*, 498(1):1420–1439, October 2020. doi: 10.1093/mnras/stz3094.

A. Appendix

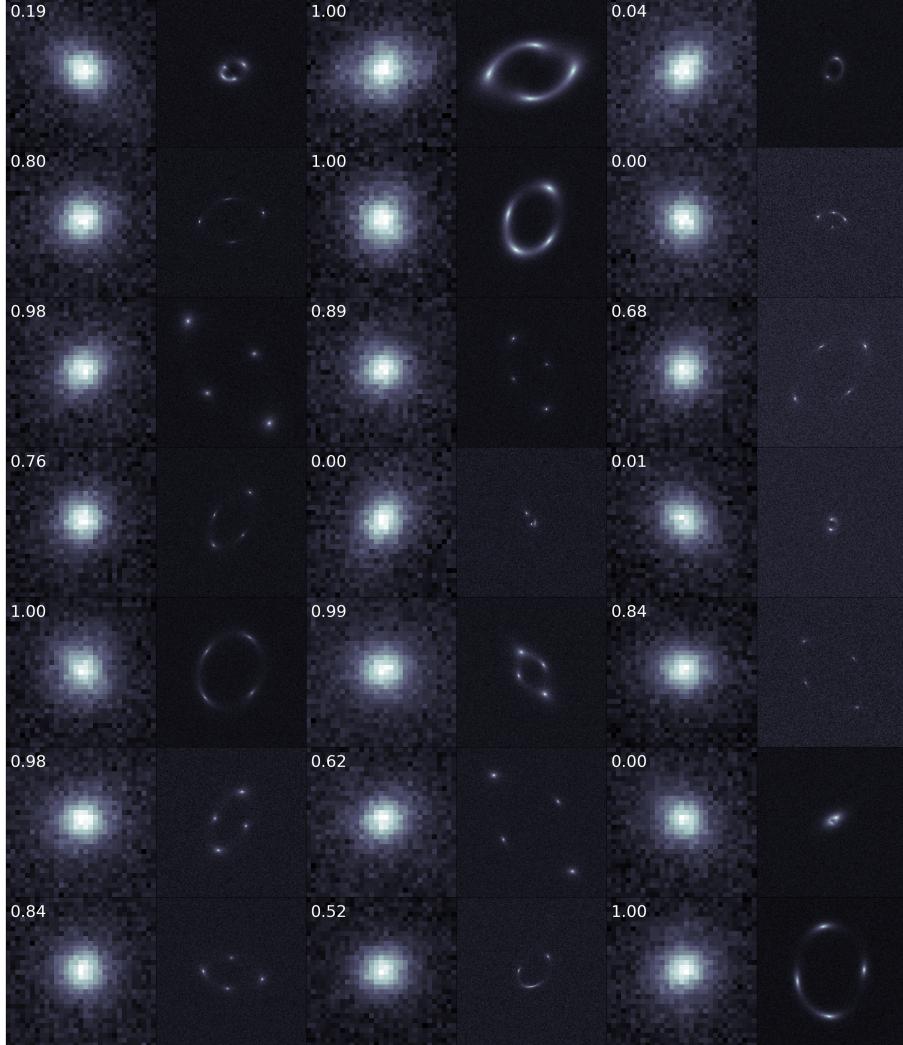


Figure 4. Examples of generated strong gravitational lensing simulations. The set of simulations used consists of low-resolution wide-area sky surveys (left) and the corresponding follow-up high-resolution image (right). The probability of detection p_{det} predicted by our neural network model is shown in the top corner of each low-resolution image. The strong lensing simulations are composed of a Singular Isothermal Ellipsoid (Kormann et al., 1994, SIE) with added external shear representing the main lens deflector and a background source following a Sérsic brightness profile (Sérsic, 1963). We also add stellar light following a Sérsic profile to the main lens deflector galaxy in the low-resolution images. This is not done for the high-resolution follow-up images as we assume that the stellar light from the main deflector galaxy is removed in a data preprocessing stage. As a whole, our simulations can be fully described by a 13-dimensional parameter vector θ .

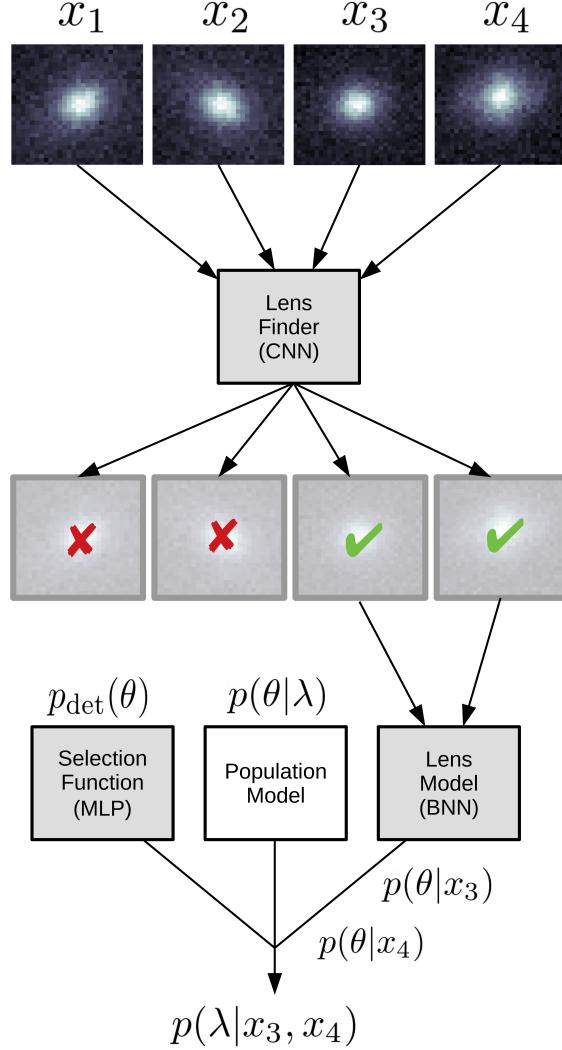


Figure 5. Flow chart illustrating the general setup for the inference of the population hyperparameter posterior $p(\lambda|\{x_i\})$ for an example with four generated strong lensing observations $\{x_1, x_2, x_3, x_4\}$. The grey colored boxes are the neural networks used to perform the hierarchical inference task. The lens finder (CNN-based detector) is applied to the observations and classifies whether they contain strong lensing. For the images detected as strong lenses, the posterior of lensing model parameters $p(\theta|x_i)$ is obtained by applying a Bayesian Neural Network (BNN) on the followup high-resolution images. The population distribution $p(\theta|\lambda)$, neural selection function $p_{\text{det}}(\theta)$ and predicted lensing posteriors (in this example, $p(\theta|x_3)$ and $p(\theta|x_4)$) are used to compute the posterior of population-level parameters $p(\lambda|x_3, x_4)$.