arXiv:1911.03867v1 [astro-ph.IM] 10 Nov 2019

Modular Deep Learning Analysis of Galaxy-Scale Strong Lensing Images

Sandeep Madireddy[1], **Nan Li**[2], **Nesar Ramachandra**[1],
Prasanna Balaprakash[1], **Salman Habib**[1]
[1]Argonne National Laboratory, [2] University of Nottingham
{smadireddy, nramachandra, pbalapra, habib}@anl.gov, nan.li@nottingham.ac.uk

Abstract

Strong gravitational lensing of astrophysical sources by foreground galaxies is a powerful cosmological tool. While such lens systems are relatively rare in the Universe, the number of detectable galaxy-scale strong lenses is expected to grow dramatically with next-generation optical surveys, numbering in the hundreds of thousands, out of tens of billions of candidate images. Automated and efficient approaches will be necessary in order to find and analyze these strong lens systems. To this end, we implement a novel, modular, end-to-end deep learning pipeline for denoising, deblending, searching, and modeling galaxy-galaxy strong lenses (GGSLs). To train and quantify the performance of our pipeline, we create a dataset of 1 million synthetic strong lensing images using state-of-the-art simulations for next-generation sky surveys. When these pretrained modules were used as a pipeline for inference, we found that the classification (searching GGSL) accuracy improved significantly—from 82% with the baseline to 90%, while the regression (modeling GGSL) accuracy improved by 25% over the baseline.

1 Introduction

Gravitational lensing is the deflection of light rays as they traverse through the curved space caused by the presence of mass. In the present era of precision cosmology, gravitational lensing has become a powerful probe in many areas of astrophysics and cosmology, from stellar scale to cosmological scale. Galaxy-galaxy strong lensing (GGSL) is a particular case of gravitational lensing in which the background source and foreground lens are both galaxies and the lensing system is sufficient to distort images of sources into arcs or even rings, depending on the relative angular position of the two objects. Since the discovery of the first GGSL system in 1988 [13], many valuable scientific applications have been realized, such as studying galaxy mass density profiles [32, 30, 19], detecting and inferring galaxy substructure [35, 14, 2, 4], measuring cosmological parameters [6, 28, 33], investigating the nature of high-redshift galaxies [3, 9, 29], and constraining the properties of self-interacting dark matter candidates [31, 11, 18]

With the capabilities of the next-generation telescopes such as the Large Synoptic Survey Telescope (LSST),[1], Euclid[2] the number of known GGSLs is predicted to increase by several orders of magnitude [5]. The strong gravitational lens finding challenge project [24] proved the success of applying machine learning approaches to detect GGSL systems in an automated manner. Lanusse et al. [20], Morningstar et al. [25], Hezaveh et al. [15], Levasseur et al. [21] and Pearson et al. [27] have shown the feasibility and reliability of utilizing deep learning to model strong lenses as a vastly more efficient alternative to traditional parametric methods. Fast forward modeling for strong lensing image reconstructions [26] may also be combined with inference pipelines such as Markov chain Monte Carlo for lensing parameter estimation. However, the preprocessing of the original images—for example, deblending and denoising with machine learning—is still in its infancy.

In this paper, we address this growing need for preparing automated analysis for GGSLs in two ways. First, we create a dataset of 1 million simulated images (500K GGSLs and 500K non-GGSLs) using a catalog of GGSLs and a state-of-the-art semi-analytic catalog of galaxies named cosmoDC2 into a strong lensing simulation program named PICS. To demonstrate the feasibility of the pipeline for analyzing GGSLs, we use only 120K simulated images (60K GGSLs and 60K non-GGSLs) out of the 1 million images. However, we use the 1 million images to quantify the performance of our

[1] `https://www.lsst.org/` [2] `https://www.euclid-ec.org/`

Second Workshop on Machine Learning and the Physical Sciences (NeurIPS 2019), Vancouver, Canada.

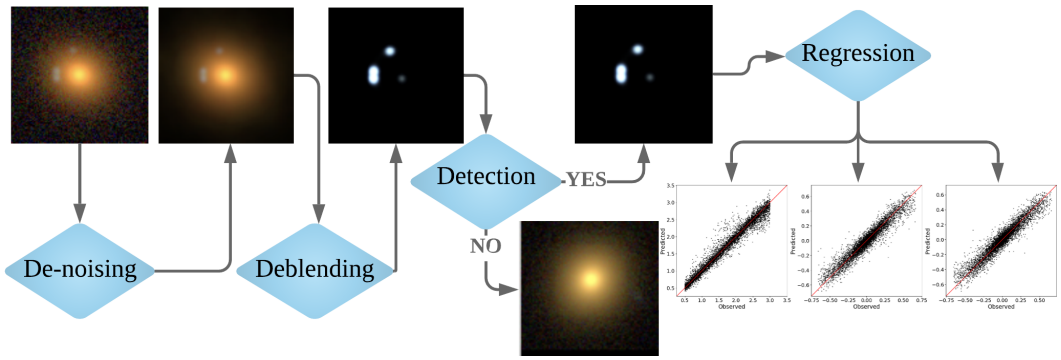


Figure 1: Machine learning pipeline for analysis of galaxy-scale strong lensed systems.

pipeline in further studies. Second, we develop an end-to-end machine learning pipeline for automated lens finding and characterization for GGSLs, which consists of four modules—denoising, deblending, lens identification, and lens characterization. We adopt Deep Residual Networks (ResNet)-based fully convolutional neural network architectures for denoising the original pixelized images and removing the lens light in the deblending module. The lens identification and characterization modules perform classification and regression, respectively, which are both built by using ResNet-50 architecture. We demonstrate considerable improvement over lens finding and characterization without the pipeline, and we discuss potential avenues for future improvement.

2 Data Preparation – Simulations

We created a realistic simulated dataset, including 500K GGSLs and 500K non-GGSLs by adopting a catalog of strong lenses [5] (hereafter, Collett15) and a state-of-the-art extragalactic catalog [17]. Collett15 provides the mass models and simple light models of both lens and source galaxies; on the other hand, cosmoDC2 provides more realistic light profiles of galaxies containing bulges and disks. To create the inputs for our strong lensing simulation program named PICS [22] by connecting the mass profiles from Collett15 and light profiles from cosmoDC2, we crossmatch the apparent magnitudes, axis ratios, position angles, and redshifts of the galaxies from Collett15 and CosmoDC2.

The mass model of an individual lens galaxy is a singular-isothermal ellipsoid (SIE) as is adopted in Collett15, which not only is analytically tractable but also has been found to be consistent with models of individual lenses and lens statistics on the length scales relevant for strong lensing [16, 10, 8]. Accordingly, the deflection maps can be given by the parameters of the positions, velocity dispersion, axis ratio, position angle, and redshift of lens as well as redshifts of source galaxies, i,e, $\{x_1, x_2, \sigma_v, q_l, \phi_l, z_l, z_s\}$. Since $\{x_1, x_2\}$ can be fixed to $\{0, 0\}$ by centering the cutouts at the lens galaxies and since the lensing strength (i.e., Einstein radius) can be given by $R_{\rm ein} = 4\pi(\sigma_v/c)^2 D(z_l, z_s)/D(z_s)$, the parameter array can be simplified as $\{R_{\rm ein}, q_l, \phi_l\}$, where c is the speed of light, $D(z_l, z_s)$ and $D(z_s)$ are the angular diameter distance from the deflector to the source and from the observer to the source, respectively.

We added noise and the point spread function (PSF) to make the images realistic using models of a ground-based-like telescope from [5, 7]. The noise model is a mix between read noise, which is a Gaussian-like noise, and shot noise, which is a Poisson-like noise that can be calculated according to the flux in the pixelized images. The PSF model is also a Gaussian function with different full width at half maximum (FWHM) of $[g, r, i]$ bands. Examples are shown in Appendix A, Fig. 2. The nonlensing systems are generated in the same way but with the strong lensing effects removed by considering the deflection angles as zeros.

3 Methodology – Pipeline Training and Inference

Our proposed machine learning pipeline consists of four modules—denoising, source separation (deblending), lens searching (classification), and Lens modeling (regression)—as shown in Fig. 1

Denoising is an image restoration approach in which the goal is to recover a clean image X from a noisy observation Y. Traditionally, image denoising has been posed as an inverse problem, where optimization approaches and special purpose regularizers (known as image priors) have been used to achieve this [1]. Recently, deep-learning-based approaches have been being increasingly adopted and are currently the state-of-the-art algorithms [23, 36] for image denoising. We adopt an enhanced deep super-residual network (EDSR) architecture [23] that was proposed for a specific type of image restoration known as super resolution. The residual network (ResNet [12]) incorporates skip connections between residual blocks (which consist of convolution, batch normalization, and

nonlinear activation layers) in a deep network, and has been shown to work well for a variety of tasks [34]. The residual networks overcome the vanishing gradient problem by learning the function mapping for the residual (with respect to inputs). EDSR proposes to get rid of the batch normalization layers that are deemed unnecessary for the image-to-image tasks. Since the inputs and the outputs for denoising have the same resolution, we removed the up-sampling layer from the EDSR architecture which is composed of 16 residual blocks each containing two convolutional layers and an ReLU nonlinear activation function. The convolution layers use $3x3$ kernels and 256 feature channels. The *source separation (deblending)* module decouples the lensed light and the source galaxy from the observations. The module also utilizes the same modified EDSR architecture that was used for the denoising module. The reason is that the source separation is also an image-to-image task that takes the images with coupled source and foreground galaxies as input and outputs the corresponding lensed or nonlensed source galaxy that is separated from the foreground lens. The *classification* module is used to detect the lensing systems from the source separated images. In other words, each of the observed image needs to be classified as to whether it is a lensed or a nonlensed system. We utilize the ResNet-50 architecture to perform this classification. In this architecture, each residual block is three layers deep and consists of convolution layers with channel size increasing from 64 to 2048 and the filter size being either 1×1 or 3×3. The *parameter estimation (regression)* module takes the source-separated galaxy and predict its characteristics—Einstein radius, axis ratio, and position angle. The parameter estimation also uses the ResNet-50 architecture that has been adopted for the classification, but the last layer is replaced with a fully connected layer that is used to predict the three continuous quantities. We also considered the case where we single model was used for denoising and debleding together, but we found the discussed pipeline (Figure 1) to be the best and hence do not discuss the former in detail, in the interest of space.

4 Results and Discussion

Each of the four modules—denoising, deblending, classification, and regression—was trained individually using the corresponding parts of the simulation data. Once the model was trained, the weights in these models were fixed and deployed as part of the inference pipeline, where the predictions from each module were fed into the subsequent module with the end goal of characterizing the lensed galaxies. The results for training the modules will be discussed first, followed by inference.

4.1 Training

The denoising ESDR model was trained for $12,50$ epochs using $5,000$ images (and tested on 200), where the inputs were noisy and blended galaxy images (`Noisy-Sim`) and the ground truth was the corresponding noiseless blended images (`Noiseless-Sim`) from the simulation. We used the peak signal-to-noise ratio (PSNR) to evaluate the denoising accuracy.

The accuracy metrics corresponding on the test data using the trained denoising model are shown in Table 1a in Appendix A. First, the difference between `Noisy-Sim` and `Noiseless-Sim` is shown to demonstrate the effect of the noise. The mean value for PSNR over all the test data is 22.56, indicating that in fact the noise has a significant effect on the image. Then, the ability of the denoising machine learning model to predict the denoised images from the noisy images is measured by comparing the model prediction (`Noiseless-ML`) with the corresponding ground truth (`Noiseless-Sim`). The metric value of 47.55 for PSNR indicates very good noise removal by the trained model. To measure the accuracy of the deblending module, we first compared the `Noiseless-Sim` with noiseless and deblended simulation data (`Deblended-Sim`) to characterise the difference between the input and the corresponding ground truth that the prediction seeks to match (Table 1a). The mean value over all the test data for PSNR is 14.47, which indicates a significant difference between these image pairs. Then, the output from the deblending machine learning model (`Deblended-ML`) was compared with the ground truth (`Deblended-Sim`) on the test data to characterize the predictive accuracy. The PSNR value of 35.03 indicates a good recovery of the source galaxy by deblending.

The classification module was trained over 108,000 images with a batch size of 256, 15 epochs, and a learning rate that decays by half every two epochs starting with a value of $1e-4$ with the Adam optimizer. The mean classification accuracy (over two classes) was used to measure the accuracy of the classification model (Table 1b). As a baseline we trained a classification model to predict the label directly from the noisy blended simulation images (`Noisy-Sim`) and evaluated the metrics on the same $12,000$ corresponding test data images. The mean accuracy was found to be 0.82, while the classification model trained with `Deblended-Sim` gave a mean accuracy of 0.99, which is a significant improvement overthe baseline.

For the parameter estimation (regression) module, we used the same ResNet-50 architecture but with the last layer being a fully connected one to predict the three continuous parameters. Only the lensed images in the deblended simulation data `Deblended-Sim-Len` were used to train the

regression model. Hence, a total of $54,000$ were used for training and $6,000$ for testing. The same batch size and learning rate schedule used for classification were employed for regression as well, while the number of epochs was increased to 30. The regression accuracy was measured by using the mean absolute error (MAE) in the normalized ([0,1] w.r.t the maximum and minimum of training data) coordinates, as shown in Table 1b; the plots comparing the observed and predicted are shown in Fig. 3. The regression accuracy for training is 0.01 for MAE, which indicates a very good agreement with the ground truth. For the test data, the corresponding MAE is 0.03, while the baseline MAE is 0.08.

4.2 Inference

With the inference pipeline, we looked at an application scenario where all four modules were used in unison to predict the galaxy parameters from the noisy observations. The input to the denoising module was the full 12000 `Noisy-Sim` data, and we evaluated the denoising performance of the inference data using a procedure similar to that used for the test data, where the similarity between `Noisy-Sim` data and the `Noiseless-Sim` was calculated with the three metrics. We found the results to be similar to those obtained on the test data, giving us confidence that there wss no significant change in the noise distribution. Next, we compared the performance of the predictions of the deblending model (`Noiseless-ML`) with the ground truth (`Noiseless-Sim`) and again found the metrics to be close to those obtained for test data, thus validating the predictive capability and generalizability of this denoising model beyond the data it's trained on. For the deblending process, the denoised model predictions (`Noiseless-ML`) were taken as input and the corresponding deblended outputs were obtained (`Deblended-ML-ML`) as output. The accuracy was evaluated with respect to the ground truth deblended images `Deblended-Sim` for `Noiseless-ML` and `Deblended-ML-ML` over all the 120000 images (Table 1c). We found the PSNR for the latter to be 26.87, which is lower than that for the test data (in the training phase) but is significantly better than the baseline of 13.59.

For the classification inference, we calculated the mean accuracy for the deblending scenario and found that the accuracy is lower than for the test data (in the training phase), with a mean accuracy of .90 compared with 0.99. However, this accuracy is much higher than the baseline case accuracy of 0.82.

For the regression inference, we calculated the MAE for the deblending scenario and found that their regression accuracy (Table 1d) is slightly lower (MAE of 0.06) than the accuracies obtained on the test data but is an improvement over the baseline of 0.08.

Limitations of the Denoising/Deblending Modules: Although we obtained good training and test accuracy for all the training modules and significant improvement in the lens finding (classification) over the baseline for the inference pipeline, the lens characterization accuracy improvement in inference is only marginal. We attribute this to two factors: (1) sensitivity of the deblending module to the denoising input, where we found that even though the PSNR is close to ideal, the minor differences with the ground truth ($< 0.01\%$ cause additional features in the deblended image (Fig. 2(e) in Appendix B) for some cases; and (2) the processes of denoising and deblending, which work well for extracting bright lensed arcs but can erase the faint counter-images of the primary lensed images because of the high contrasts between the images of the lenses and the counterimage. These factors may bias the ellipticity and inner density slopes of lens galaxies. We will try to improve these issues by involving larger training sets and more complex models in the follow-up work.

5 Conclusions

Combining high-fidelity simulation data and a systematic machine learning pipeline is crucial for developing fast and accurate GSSL analysis techniques for future cosmological surveys. To this end, we proposed a dataset of 1 million synthetic images (500k GGSLs and 500k non-GGSLs), which is the largest simulation for GGSL ever made, and we developed an end-to-end machine learning pipeline with separate modules for denoising, deblending, lens searching, and lens modeling that is trained on this data. We demonstrate good denoising and deblending performance on both the training and inference (compared with the ground truth) and, consequently, a significant improvement in classification and regression over the baseline (working directly with the noisy blended data). We also identify a few limitations of the simulation data, which lead to underestimated contamination from substructures in the context of both mass and light profiles of galaxies and in the denoising/deblending model, which either misses counterimages or introduces additional artifacts. We plan to address these issues by scaling up to the full million image dataset, training the denoising and deblending model together, and employing a hyperparameter search to improve the classification and regression accuracies. In addition, we will explore uncertainty quantification for the lens modeling output through probabilistic regression. Eventually, the pipeline intended to be used for real-time lens

finding and characterization with data from next-generation large-scale sky surveys such as Euclid, LSST, and WFIRST.

Acknowledgments

This material is based upon work supported by the U.S. Department of Energy (DOE), Office of Science, Office of Advanced Scientific Computing Research, under Contract DE-AC02-06CH11357. We gratefully acknowledge the computing resources provided and operated by the Joint Laboratory for System Evaluation (JLSE) at Argonne National Laboratory. The work is also supported by the UK Science and Technology Facilities Council (STFC).

References

[1] Saeed Anwar, Salman Khan, and Nick Barnes. A deep journey into super-resolution: A survey. *arXiv preprint arXiv:1904.07523*, 2019.

[2] 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. *ArXiv e-prints*, 2018.

[3] M. B. Bayliss, K. Sharon, A. Acharyya, M. D. Gladders, J. R. Rigby, F. Bian, R. Bordoloi, J. Runnoe, H. Dahle, L. Kewley, M. Florian, T. Johnson, and R. Paterno-Mahler. Spatially Resolved Patchy Lyα Emission within the Central Kiloparsec of a Strongly Lensed Quasar Host Galaxy at z = 2.8. *The Astrophysical Journal Letters*, 845:L14, 2017.

[4] 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. *arXiv e-prints*, page arXiv:1909.02005, Sept. 2019.

[5] T. E. Collett. The Population of Galaxy-Galaxy Strong Lenses in Forthcoming Optical Imaging Surveys. *The Astrophysical Journal*, 811:20, September 2015.

[6] T. E. Collett and M. W. Auger. Cosmological constraints from the double source plane lens SDSSJ0946+1006. *Monthly Notices of the Royal Astronomical Society*, 443:969–976, 2014.

[7] A. J. Connolly, John Peterson, J. Garrett Jernigan, Robert Abel, Justin Bankert, Chihway Chang, Charles F. Claver, Robert Gibson, David K. Gilmore, Emily Grace, R. Lynne Jones, Zeljko Ivezic, James Jee, Mario Juric, Steven M. Kahn, Victor L. Krabbendam, Simon Krughoff, Suzanne Lorenz, James Pizagno, Andrew Rasmussen, Nathan Todd, J. Anthony Tyson, and Mallory Young. Simulating the lsst system. *Modeling, Systems Engineering, and Project Management for Astronomy IV*, 7738, 2010.

[8] S. Dye, N. W. Evans, V. Belokurov, S. J. Warren, and P. Hewett. Models of the Cosmic Horseshoe gravitational lens J1004+4112. *Monthly Notices of the Royal Astronomical Society*, 388(1):384–392, July 2008.

[9] S. Dye, C. Furlanetto, L. Dunne, S. A. Eales, M. Negrello, H. Nayyeri, P. P. van der Werf, S. Serjeant, D. Farrah, M. J. Michałowski, M. Baes, L. Marchetti, A. Cooray, D. A. Riechers, and A. Amvrosiadis. Modelling high-resolution ALMA observations of strongly lensed highly star-forming galaxies detected by Herschel. *Monthly Notices of the Royal Astronomical Society*, 476:4383–4394, 2018.

[10] Raphaël Gavazzi, Tommaso Treu, Jason D. Rhodes, Léon V. E. Koopmans, Adam S. Bolton, Scott Burles, Richard J. Massey, and Leonidas A. Moustakas. The Sloan Lens ACS Survey, IV: the mass density profile of early-type galaxies out to 100 effective radii. *The Astrophysical Journal*, 667(1):176–190, Sept. 2007.

[11] D. Gilman, S. Birrer, T. Treu, and C. R. Keeton. Probing the nature of dark matter by forward modeling flux ratios in strong gravitational lenses. *ArXiv e-prints*, 2017.

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

[13] J. N. Hewitt, E. L. Turner, D. P. Schneider, B. F. Burke, and G. I. Langston. Unusual radio source MG1131+0456 - A possible Einstein ring. *Nature*, 333:537–540, 1988.

[14] Y. D. Hezaveh, N. Dalal, D. P. Marrone, Y.-Y. Mao, W. Morningstar, D. Wen, R. D. Blandford, J. E. Carlstrom, C. D. Fassnacht, G. P. Holder, A. Kemball, P. J. Marshall, N. Murray, L. Perreault Levasseur, J. D. Vieira, and R. H. Wechsler. Detection of lensing substructure using ALMA observations of the Dusty Galaxy SDP.81. *The Astrophysical Journal*, 20.

[15] Y. D. Hezaveh, L. P. Levasseur, and P. J. Marshall. Fast automated analysis of strong gravitational lenses with convolutional neural networks. *Nature*, 548:555–557, 2017.

[16] Léon V. E. Koopmans, Tommaso Treu, Adam S. Bolton, Scott Burles, and Leonidas A. Moustakas. The Sloan Lens ACS Survey, III: The structure and formation of early-type galaxies and their evolution since z $\sim$1. *The Astrophysical Journal*, 649(2):599–615, Oct. 2006.

[17] Danila Korytov, Andrew Hearin, Eve Kovacs, Patricia Larsen, Esteban Rangel, Joseph Hollowed, Andrew J. Benson, Katrin Heitmann, Yao-Yuan Mao, Anita Bahmanyar, Chihway Chang, Duncan Campbell, Joseph Derose, Hal Finkel, Nicholas Frontiere, Eric Gawiser, Salman Habib, Benjamin Joachimi, François Lanusse, Nan Li, Rachel Mandelbaum, Christopher Morrison, Jeffrey A. Newman, Adrian Pope, Eli Rykoff, Melanie Simet, Chun-Hao To, Vinu Vikraman, Risa H. Wechsler, and Martin White. CosmoDC2: A Synthetic Sky Catalog for Dark Energy Science with LSST. *arXiv e-prints*, page arXiv:1907.06530, July 2019.

[18] J. Kummer, F. Kahlhoefer, and K. Schmidt-Hoberg. Effective description of dark matter self-interactions in small dark matter haloes. *Monthly Notices of the Royal Astronomical Society*, 474:388–399, 2018.

[19] R. Küng, P. Saha, I. Ferreras, E. Baeten, J. Coles, C. Cornen, C. Macmillan, P. Marshall, A. More, L. Oswald, A. Verma, and J. K. Wilcox. Models of gravitational lens candidates from Space Warps CFHTLS. *Monthly Notices of the Royal Astronomical Society*, 474:3700–3713, March 2018.

[20] F. Lanusse, Q. Ma, N. Li, T. E. Collett, C.-L. Li, S. Ravanbakhsh, R. Mandelbaum, and B. Póczos. CMU DeepLens: deep learning for automatic image-based galaxy-galaxy strong lens finding. *Monthly Notices of the Royal Astronomical Society*, 473:3895–3906, 2018.

[21] Laurence Perreault Levasseur, Yashar D Hezaveh, and Risa H Wechsler. Uncertainties in parameters estimated with neural networks: Application to strong gravitational lensing. *arXiv preprint arXiv:1708.08843*, 2017.

[22] Nan Li, Michael D. Gladders, Esteban M. Rangel, Michael K. Florian, Lindsey E. Bleem, Katrin Heitmann, Salman Habib, and Patricia Fasel. PICS: Simulations of strong gravitational lensing in galaxy clusters. *The Astrophysical Journal*, 828(1):54, Sept. 2016.

[23] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 136–144, 2017.

[24] R. B. Metcalf, M. Meneghetti, C. Avestruz, F. Bellagamba, C. R. Bom, E. Bertin, R. Cabanac, F. Courbin, A. Davies, E. Decencière, R. Flamary, R. Gavazzi, M. Geiger, P. Hartley, M. Huertas-Company, N. Jackson, C. Jacobs, E. Jullo, J.-P. Kneib, L. V. E. Koopmans, F. Lanusse, C.-L. Li, Q. Ma, M. Makler, N. Li, M. Lightman, C. E. Petrillo, S. Serjeant, C. Schäfer, A. Sonnenfeld, A. Tagore, C. Tortora, D. Tuccillo, M. B. Valentín, S. Velasco-Forero, G. A. Verdoes Kleijn, and G. Vernardos. The strong gravitational lens finding challenge. *Astronomy & Astrophysics*, 625:A119, May 2019.

[25] 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. *arXiv preprint arXiv:1808.00011*, 2018.

[26] Warren R Morningstar, Laurence Perreault Levasseur, Yashar D Hezaveh, Roger Blandford, Phil Marshall, Patrick Putzky, Thomas D Rueter, Risa Wechsler, and Max Welling. Data-driven reconstruction of gravitationally lensed galaxies using recurrent inference machines. *arXiv preprint arXiv:1901.01359*, 2019.

[27] J. Pearson, N. Li, and S. Dye. The use of convolutional neural networks for modelling large optically-selected strong galaxy-lens samples. *Monthly Notices of the Royal Astronomical Society*, 488:991–1004, 2019.

[28] A. Rana, D. Jain, S. Mahajan, A. Mukherjee, and R. F. L. Holanda. Probing the cosmic distance duality relation using time delay lenses. *Journal of Cosmology and Astroparticle Physics*, 7:010, 2017.

[29] P. Sharda, C. Federrath, E. daÂ Cunha, A. M. Swinbank, and S. Dye. Testing star formation laws in a starburst galaxy at redshift 3 resolved with ALMA. *Monthly Notices of the Royal Astronomical Society*, 477:4380–4390, 2018.

[30] Y. Shu, A. S. Bolton, S. Mao, C. S. Kochanek, I. Pérez-Fournon, M. Oguri, A. D. Montero-Dorta, M. A. Cornachione, R. Marques-Chaves, Z. Zheng, J. R. Brownstein, and B. Ménard. The BOSS Emission-line Lens Survey. IV. Smooth Lens Models for the BELLS GALLERY Sample. *The Astrophysical Journal*, 833:264, 2016.

[31] Y. Shu, A. S. Bolton, L. A. Moustakas, D. Stern, A. Dey, J. R. Brownstein, S. Burles, and H. Spinrad. Kiloparsec Mass/Light Offsets in the Galaxy Pair-Lyα Emitter Lens System SDSS J1011+0143. *The Astrophysical Journal*, 820:43, 2016.

[32] A. Sonnenfeld, T. Treu, P. J. Marshall, S. H. Suyu, R. Gavazzi, M. W. Auger, and C. Nipoti. The SL2S Galaxy-scale Lens Sample. V. Dark Matter Halos and Stellar IMF of Massive Early-type Galaxies Out to Redshift 0.8. *The Astrophysical Journal*, 800:94, 2015.

[33] S. H. Suyu, V. Bonvin, F. Courbin, C. D. Fassnacht, C. E. Rusu, D. Sluse, T. Treu, K. C. Wong, M. W. Auger, X. Ding, S. Hilbert, P. J. Marshall, N. Rumbaugh, A. Sonnenfeld, M. Tewes, O. Tihhonova, A. Agnello, R. D. Blandford, G. C.-F. Chen, T. Collett, L. V. E. Koopmans, K. Liao, G. Meylan, and C. Spiniello. H0LiCOW - I. H_0 Lenses in COSMOGRAIL's Wellspring: program overview. *Monthly Notices of the Royal Astronomical Society*, 468:2590–2604, July 2017.

[34] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A Alemi. Inception-v4, Inception-ResNet and the impact of residual connections on learning. In *AAAI*, volume 4, page 12, 2017.

[35] S. Vegetti, L. V. E. Koopmans, M. W. Auger, T. Treu, and A. S. Bolton. Inference of the cold dark matter substructure mass function at z = 0.2 using strong gravitational lenses. *Monthly Notices of the Royal Astronomical Society*, 442:2017–2035, 2014.

[36] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2472–2481, 2018.

A Supplemental Tables and Figures

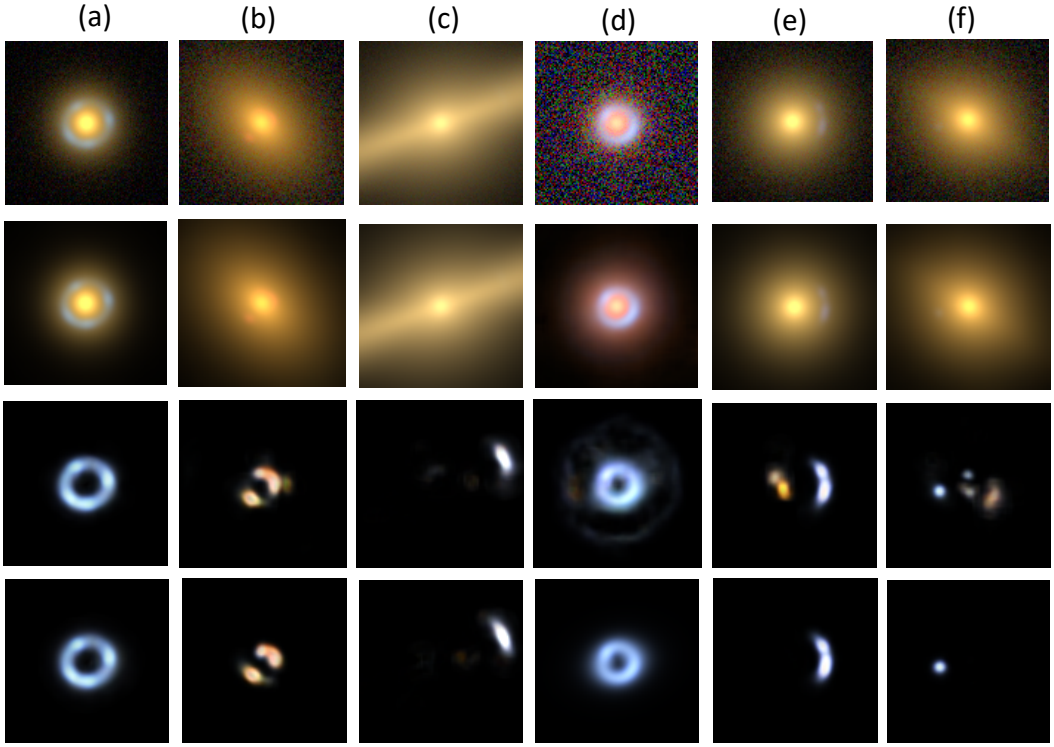


Figure 2: First row: `Noisy-Sim` data from the simulation. Second row: `Noiseless-ML` output from the denoising module at inference. Third row: `Deblended-ML-ML` output from the deblending model with the denoised model input at inference. Fourth row:`Deblended-Sim-ML` output from the deblending model with the denoised simulation input (`Noiseless-Sim`) at inference.

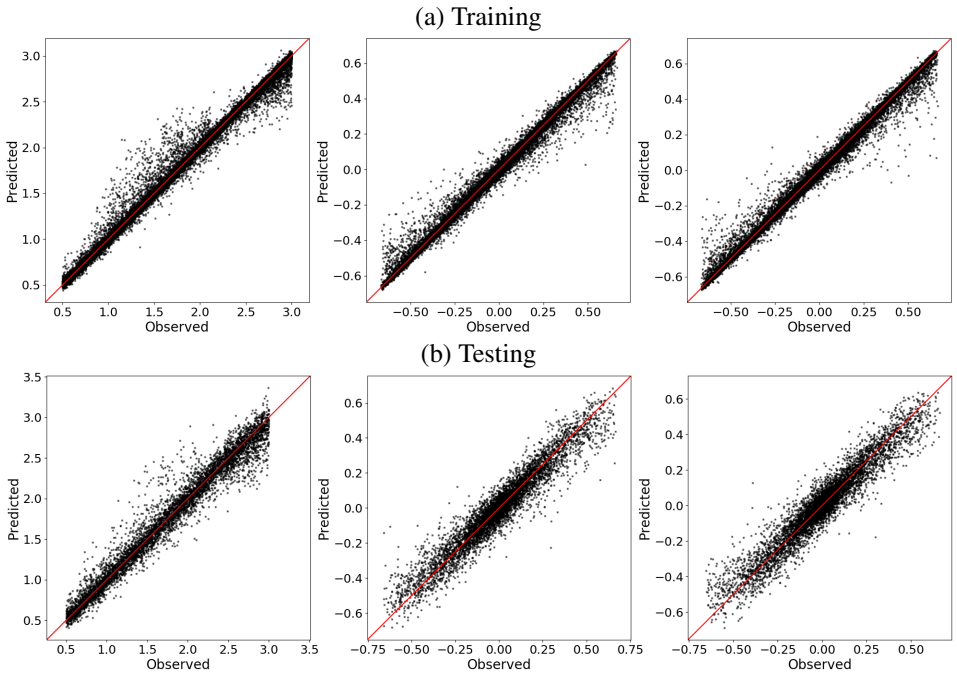


Figure 3: Comparison of the observed (`Noiseless-Sim`) and predicted (`Noiseless-ML`) data corresponding to training and testing data for lens characterization (regression) during the training phase.

Table 1: Accuracy metrics for training and inference with the machine learning pipeline.

(a) Training – PSNR metrics for denoising and deblending on 200 images

Training	**PSNR**
Denoising	
Noisy-Sim	22.56
Noiseless-ML	47.91
Deblending	
Noiseless-Sim	14.47
Deblended-ML	35.03

(b) Training – classification and regression metrics on $12,000$ images

Training	**Mean acc**
Classification	
Noisy-Sim	0.82
Noiseless-Sim	0.99
	MAE
Regression	
Noisy-blended-Sim	0.08
Noiseless-Sim-Len	0.03

(c) Inference - PSNR metrics for Denoising and Deblending on $120,000$ images

Inference	**PSNR**
Denoising	
Noisy-Sim	22.71
Noiseless-ML	47.15
Deblending	
Noiseless-ML	13.59
Deblended-ML-ML	26.87

(d) Inference – classification and regression metrics on $12,000$ images

Inference	**Mean acc**
Classification	
Deblended-ML-ML	0.90
	MAE
Regression	
Deblended-ML-ML	0.06

B Limitations of the Simulation Model

We adopt SIE as the mass model of the lenses, which is insufficient for studying the influence of the subtle structures of the lens galaxies on the performance of analyzing strong lenses using our pipeline. To make the simulation more realistic, we plan to adopt the particle data from cosmological N-body simulations and stellar mass distributions from semi-analytical models for presenting the mass distribution of lens galaxies. Furthermore, for image simulation, we involve the images of sources and lenses only; but in real observations, images of galaxies on the line of sight are also considerable, and the cosmoDC2 light cone is helpful to include these effects. The study is focused on ground-based-like telescopes such as LSST, so the light profiles with bulges and disks are sufficient because of the coarse pixelization and large PSF. However, for the case of space-based-like telescopes such as Euclid and WFIRST, the light profile lacks detailed structures of galaxies such as spirals and clumps. We are attempting to attach substructures onto the galaxies by using GANs in a parallel project. Another issue is the overkill problem in the processes of denoising and deblending; that is, faint counterimages of the primary lensed images can be erased on some level because of the high contrasts between the images of the lenses and the counterimages. We will try to avoid potential biases because of this problem by using more advanced algorithms in our follow-up work.

The submitted manuscript has been created by UChicago Argonne, LLC, Operator of Argonne National Laboratory ("Argonne"). Argonne, a U.S. Department of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357. The U.S. Government retains for itself, and others acting on its behalf, a paid-up nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan. http://energy.gov/downloads/doe-public-access-plan