SPEKTRA: Jurnal Fisika dan Aplikasinya
Volume 10, Issue 3, December 2025
Homepage: doi.org/10.21009/SPEKTRA

DOI: doi.org/10.21009/SPEKTRA.103.05

Gravitational Lens Parameters Estimation at Intermediate Redshifts Using Convolutional Neural Networks

Muhammad Doni Setiawan[1,*], Anton Timur Jaelani[2,3,4]

[1]*Astronomy Study Program, FMIPA, Institut Teknologi Bandung, Jl. Ganesa 10, Bandung 40132, Indonesia*
[2]*Astronomy Research Group and Bosscha Observatory, FMIPA, Institut Teknologi Bandung, Jl. Ganesa 10, Bandung 40132, Indonesia*
[3]*U-CoE AI-VLB, Institut Teknologi Bandung, Jl. Ganesa 10, Bandung 40132, Indonesia*
[4]*University Center of Excellence for Space Science, Technology and Innovation, Institut Teknologi Bandung, Jl. Ganesa 10, Bandung 40132, Indonesia*

*Corresponding Author Email: setiawmdoni@gmail.com

Received: 19 August 2025
Revised: 14 December 2025
Accepted: 28 December 2025
Online: 30 December 2025
Published: 30 December 2025

SPEKTRA: Jurnal Fisika dan Aplikasinya
p-ISSN: 2541-3384
e-ISSN: 2541-3392

ABSTRACT

Strong gravitational lensing serves as a powerful astrophysical probe, enabling studies of dark matter, galaxy structure, and cosmological parameters. The number of strong gravitational lensing candidates at the galaxy scale is expected to reach $\mathcal{O}\sim 5$ with ongoing and future wide-field galaxy surveys. Current modeling techniques largely rely on conventional fitting methods, such as least squares or maximum likelihood using Markov Chain Monte Carlo, which despite their effectiveness, are computationally expensive and require manual inspection. This motivates the development of faster yet accurate parameter estimation techniques. In this work, we construct a representative training dataset and develop an efficient Convolutional Neural Network to estimate lens parameters: the Einstein radius, axis ratio, and position angle. We utilize data from Public Data Release 3 of the Hyper Suprime-Cam Subaru Strategic Program, selecting lens galaxies in the range $0.3 \leq z \leq 0.9$ based on the strong-lens probability distribution. We find that the choice of loss function and regularization strategy is critical. To enhance model generalization, we leverage SpatialDropout, which outperforms standard methods by addressing the spatial correlation inherent in convolutional features. Furthermore, prediction accuracy and convergence speed are strongly affected by the distribution of the training data, highlighting the importance of an appropriate loss function. Our optimized model demonstrates robust performance, achieving a Mean Absolute Error of 0.092 arcsec for the Einstein radius, providing a scalable framework for automated analysis in future wide-field surveys.

Keywords: strong gravitational lensing, convolutional neural network, lens parameter, intermediate redshift, deep learning, SpatialDropout

INTRODUCTION

Gravitational lensing is a phenomenon in which light is deflected by a gravitational field caused by the mass distribution it traverses. This deflection provides a tool for probing mass distributions, and while general lens modelling can introduce biases, strongly lensed features such as Einstein rings are particularly valuable as they trace critical curves where the enclosed mass can be determined in a largely model-independent manner [1]. Beyond its physical applications, strong gravitational lensing also acts as a natural telescope, enabling magnified observations of distant background objects. This makes lensing an essential tool in studies of galaxy evolution and in constraining cosmological parameters.

With the advent of large-scale imaging surveys, such as the Vera C. Rubin Observatory's Legacy Survey of Space and Time (LSST), the number of detected strong lensing systems is expected to grow substantially. This anticipated data volume demands automated, fast, and accurate analysis techniques. Traditional parametric modeling approaches—such as GLEE & GLAD—offer high-precision parameter estimates, but remain semi-automated and computationally expensive, often requiring several days per lens system even with partial automation [2].

In response to these limitations, various alternative approaches based on Convolutional Neural Networks (CNNs) have been developed to accelerate and scale the estimation of lens model parameters. The application of CNNs in this field initially focused on automated detection, as demonstrated by Schaefer et al. [3] and Davies et al. [4]. The maturity of these detection methods is further exemplified by recent large-scale missions such as Euclid [5], which employs deep learning to manage massive candidate identification. Alongside these detection efforts, another key application of CNNs is direct parameter regression. For instance, CNNs capable of recovering parameters of the Singular Isothermal Ellipsoid (SIE) model with high accuracy and orders-of-magnitude speed improvement were demonstrated in Hezaveh et al. [6]. Subsequent studies introduced deeper architectures like Residual Networks (ResNets) for improved regression [7], while a hybrid LSTM-FCNN was employed to predict time delays in strongly lensed Type Ia supernovae (LSNe Ia) [8]. These efforts collectively highlight the potential of CNN-based methods to serve as efficient, accurate alternatives to traditional maximum-likelihood modeling. However, few studies have specifically optimized CNN architectures to address the distinct noise characteristics and geometric properties of intermediate-redshift lenses in ground-based surveys.

In this study, we present a dedicated CNN model trained on mock images simulated using Lenstronomy [9], where lens galaxies are drawn from HSC-SSP, with redshift selection constrained to the intermediate range ($0.3 \leq z \leq 0.9$), as reported in Wong et al. [10] to match observational conditions in real lensing surveys. Redshift selection in strong lensing searches typically focuses on intermediate-redshift lenses because this regime maximizes lensing

efficiency and detectability. The network is trained to regress three key mass parameters: Einstein radius, axis ratio, and position angle, following the parameter focus in Gawade et al. [11]. Recognizing that activation functions play a crucial role by introducing the non-linearity required to learn complex patterns [12], we diverge from the approach in Gawade et al. [11] in our architectural choices. Specifically, we adopt the SiLU activation function instead of PReLU, due to SiLU's favorable gradient behavior and training stability in deeper networks. For regularization, we employ SpatialDropout [13], which preserves spatial feature correlations and addresses limitations of standard dropout in image data. Finally, to ensure robust convergence, we implement the Huber loss function. As described by Gokcesu and Gokcesu [14], this function combines the robustness of absolute loss with the strong convexity of quadratic loss, maintaining smooth gradients around small errors while limiting loss growth for large deviations.

In addition to presenting our CNN design, we also investigate the impact of activation functions and dropout strategies on the accuracy of lens parameter estimation. Our results demonstrate how architectural choices can influence performance and underscore the feasibility of deploying deep learning models that bridge robust ML architectural design with precise astrophysical analysis in upcoming surveys like LSST.

METHODS

1. Data Selection

The data used in this work are drawn from the Public Data Release 3 (PDR3) of the Hyper Suprime-Cam Subaru Strategic Program (HSC-SSP). We selected raw galaxy images to serve as potential lens galaxies. To ensure sufficient mass for producing strong lensing effects, we restricted the sample to Luminous Red Galaxies (LRGs)—massive, early-type galaxies composed primarily of old stellar populations with low star formation rates [15], making them ideal candidates for strong lensing, which is why they are commonly selected in wide-field surveys such as the Survey of Gravitationally-lensed Objects in HSC Imaging (SuGOHI) [16].

Based on prior works suggesting that strong lens systems are more commonly found at intermediate redshifts [10], we applied selection criteria combining constraints from Tanaka [17] and Jaelani et al. [16], including $0.3 \leq z \leq 0.9$, stellar mass $M_* < 10^{11} M_\odot$, specific star formation rate $sSFR < 10^{-12}$, and magnitude limits of $g < 26.0$, $r < 26.0$, and $i < 23.0$. We also required extendedness > 0.9 to ensure the object is resolved.

2. Simulation of Mock Lensing Systems

Based on the selected sample of massive early-type galaxies described in the previous section, we generated a synthetic dataset of mock lensing images. This process combines analytic lens models and simulated background sources to produce realistic gravitational lensing systems for model training (see FIGURE 1).

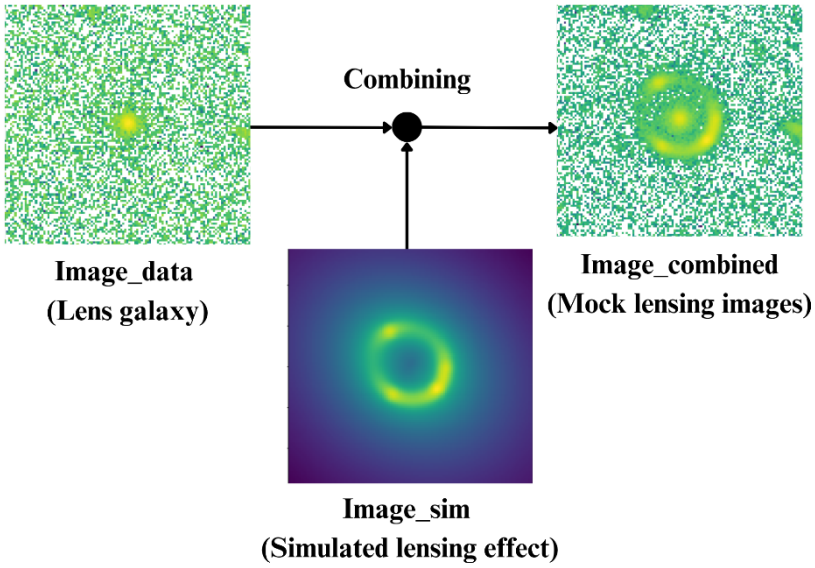

FIGURE 1. Example of mock data for a gravitational lens system. The lens galaxy comes from real HSC SSP data, while the source and lensing effect are simulated using Lenstronomy. Used for machine learning training.

These selection criteria allowed us to construct a subset of massive, morphologically suitable galaxies located within the redshift range where strong lensing is most probable. This subset was then used to generate simulated lensing images and to train the CNN model.

For the lens mass model, we adopted the Singular Isothermal Ellipsoid (SIE), one of the most commonly used analytic profiles in strong lensing studies. Compared to the Singular Isothermal Sphere (SIS), the SIE introduces ellipticity, providing a more realistic description of early-type galaxies. A conceptual overview is presented in Meneghetti [18], while the analytical formalism is given by Kormann et al. [19].

Additional components such as external shear and environmental variations were incorporated into the simulations, guided by insights from previous studies [7, 20-22]. While the lens galaxy ellipticity was not simulated but rather measured directly from real HSC data, the source galaxy ellipticity was explicitly generated as part of the mock image pipeline. These modeling choices were informed by prior literature and iteratively refined to produce realistic lensing configurations.

3. Network Architecture and Training

As previously mentioned, this study utilizes a CNN architecture. A CNN consists of hierarchically structured layers that perform gradual pattern recognition. It operates by leveraging convolutional layers to filter complex feature patterns within an image [23, 24].

Activation functions play a crucial role by introducing the non-linearity required for the neural network to learn complex patterns [25]. These functions process the weighted sum of inputs and a bias to determine whether a neuron will be activated or not [26]. Consequently, the choice of activation function significantly influences the model's performance.

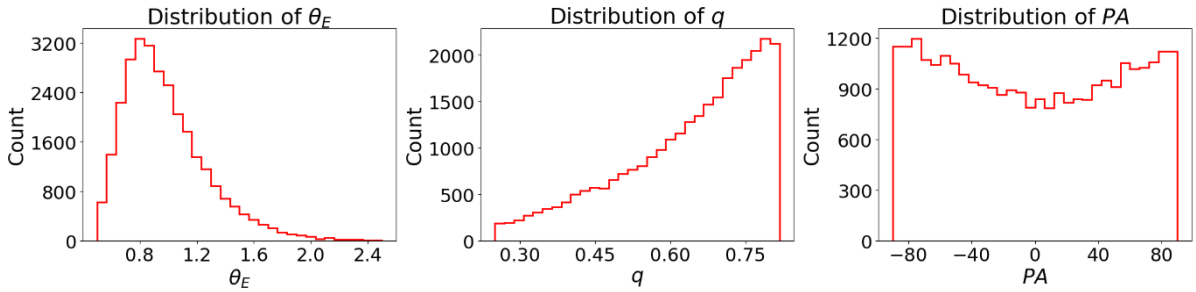

FIGURE 2. Histogram of lens parameters: from left to right are Einstein radius (θ_E), axis ratio (q), and position angle (PA). The Einstein radius distribution is skewed left (toward small values), axis ratio is skewed right (favoring rounder galaxies), and PA shows a U-shape pattern, reflecting random orientation of galaxy ellipticities.

The CNN model used in this study was specifically designed for the task of lens parameter regression. The architecture consists of four convolutional blocks, each composed of a convolutional layer, batch normalization, SiLU activation, and max pooling with a 3×3 kernel. To improve regularization, spatial dropout with a rate of 10% was applied in the third and fourth convolutional blocks.

The number of filters increases across the blocks as follows: 32, 32, 64, and 128. After the convolutional layers, a global average pooling layer is used in place of flattening to reduce overfitting and preserve spatial information. This is followed by two fully connected layers with 128 and 64 neurons, respectively, and a final dense output layer with 3 neurons corresponding to the predicted parameters: Einstein radius (θ_E), axis ratio (q), and position angle (PA). The distribution of the main parameters of the mock lens systems are shown in FIGURE 2. The input to the model consists of mock lensing images with a size of 101×101 pixels and three channels (gri-bands). The total dataset includes 22000 images, comprising both original and augmented data. Data augmentation was performed through rotations of 30°, 45°, and 90°.

For training, we used the Adam optimizer with an initial learning rate of 1×10^{-3}. The data set is split into fixed 70% for training, 20% for validation, and 10% for testing. The Huber loss function was selected due to its robustness against outliers and stability across skewed data distributions. We employed early stopping to prevent overfitting and implemented a learning rate scheduler that reduces the learning rate when the validation loss plateaus.

RESULTS AND DISCUSSIONS

1. Model Configuration and Learning Analysis

Initial experiments adopted the PReLU activation function following Gawade et al. [8]. However, the training showed signs of overfitting despite the use of standard dropout layers. To address this, we tested the SiLU and GELU activation functions, which resulted in a more stable training process and significantly reduced overfitting. This suggests that PReLU's learnable parameter may lead to overly flexible activations, causing the network to fit noise in the training data. The comparison results for PReLU, SiLU, and GELU are shown in TABLE 1.

TABLE 1. Summarizes the preliminary comparison of three activation functions evaluated in our model. Although GeLU showed competitive results, SiLU was ultimately chosen due to its balance between performance and computational efficiency.

Activation Function	Strengths	Weaknesses
PReLU	Fast convergence during training	Tends to overfit significantly
SiLU	Stable, good validation performance	Slightly worse than GeLU in some cases
GELU	Similar performance to SiLU	More computationally expensive

We also compared standard dropout with spatial dropout for regularization in convolutional layers. Spatial dropout, which randomly drops entire feature maps, was found to be more effective in reducing overfitting. In particular, applying spatial dropout with a 10% rate only in the deeper convolutional layers (with filter sizes ≥ 64) produced better validation scores than standard dropout at 20%. Applying dropout in early layers caused a noticeable drop in performance, likely due to the loss of low-level color and shape features critical for lensing structures. The comparison results are shown in TABLE 2.

Several variations of the model structure were evaluated. Adding additional convolutional blocks before or after the core architecture (with filters such as 8 or 256) did not improve performance and in some cases led to overfitting. The optimal architecture remained with four convolutional blocks and filter sizes of 32, 32, 64, and 128. We also experimented with varying kernel sizes, including a progressive 11-7-5-3 configuration. However, this offered no improvement over using uniform 3×3 kernels. The consistent 3×3 kernel with He normal initialization yielded the highest R^2 values during validation. Reducing the input size from 120×120 pixels to 101×101 improved training stability slightly, likely because it helped the network focus on the central features of the lensing image. Furthermore, we compared using 5-channel input (*grizy*) versus 3-channel input (*gri*). Interestingly, the 3-channel input outperformed the 5-channel one, suggesting that the additional bands introduced noise or redundant information that hindered model generalization.

TABLE 2. Comparison of the effects of Dropout and SpatialDropout with different dropout rates on CNN performance for gravitational lens parameter prediction. SpatialDropout with a 20% rate applies the strongest regularization but limits loss reduction and suppresses R^2 improvement. In contrast, 10% SpatialDropout and 20% Dropout yield more balanced results, with 10% SpatialDropout offering slightly better generalization when the right epoch is selected.

Dropout Type	Rate (%)	Effect on Overfitting	Notes on R^2
Dropout	20	Reduces overfitting, fairly effective	R^2 remains stable, easy to select best epoch
SpatialDropout	20	Strong regularization	R^2 growth is suppressed, requires careful tuning
SpatialDropout	10	Mild overfitting still occurs	R^2 can be high if best epoch is chosen properly

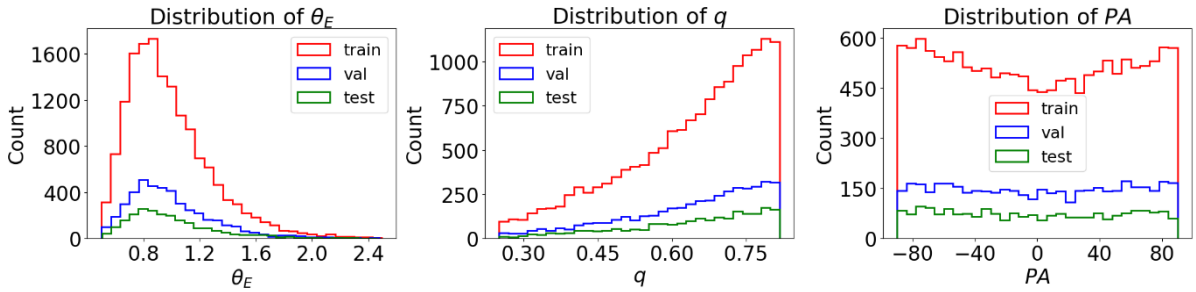

FIGURE 3. Figures from left to right are show the distributions of key lensing parameters—Einstein radius (θ_E), axis ratio (q), and position angle (PA)—for the training, validation, and test sets. Both θ_E and q exhibit skewed distributions: θ_E is left-skewed (toward lower values), while q is right-skewed (toward higher values). Meanwhile, PA shows a U-shaped distribution. This skewness motivated the use of the Huber loss function instead of MSE, due to its robustness against outliers and long-tailed regression targets.

Initial training trials used Mean Squared Error (MSE) loss with weighting schemes for different parameters. However, these approaches did not improve performance. We then analyzed the distribution of the target variables, which showed significant skewness: the θ_E was left-skewed, the q was right-skewed, and the PA had a U-shaped distribution (see FIGURE 3). The CNN is trained on a CPU i5 1335U up to 4.6 Ghz. On average, network training requires less than three hours for a training data set like the one we used.

To address the influence of outliers and imbalanced distributions, we switched to the Huber loss function with a tuned $\delta = 0.1$. This improved convergence and reduced the influence of extreme values. Additionally, we adjusted the lens mass profiles in the simulation to shift the Einstein radius distribution toward more balanced values, which helped improve R^2 scores further.

2. Prediction Results and Performance Evaluation

The final model demonstrates competitive performance across all three lensing parameters on the test set. For the θ_E, the model achieves a Mean Absolute Error (MAE) of 0.092, MSE of 0.017, and R^2 of 0.811. The q follows closely with MAE of 0.045, MSE of 0.004, and R^2 of 0.786. The highest R^2 is observed in the PA with 0.869, although its MAE of 4.991° and MSE of 373.039° due to angular discontinuity. These scores suggest that the network successfully captures the underlying relationships in the lensing images.

Scatter plots comparing predicted and true values in FIGURE 4 reveal strong agreement for θ_E and q, with points tightly clustered along the diagonal. In contrast, predictions for PA show more variance, particularly near $\pm 90°$. This behavior is likely caused by the periodic nature of angular parameters, which leads to discontinuities around the wrapping boundaries. While the model learns the overall trend well, small errors near $\pm 90°$ boundaries can result in large absolute differences. This phenomenon is also reported in Gawade et al. [8], who observed similar performance drops for PA due to angular ambiguity and modeling assumptions. In general, the position angle (PA) represents an orientation parameter. In gravitational lensing, different combinations of lens ellipticity, external shear, and source position can produce nearly indistinguishable lensed images. Rotations of the lens mass distribution can often be

compensated by corresponding changes in the shear or source morphology, resulting in degenerate solutions. Consequently, the mapping from lensed images to PA is inherently non-unique, making it challenging for machine-learning models to learn a stable and reliable relationship.

It is worth noting that the dataset used in this study was constructed to represent an idealized and relatively simple lensing configuration, yet still captures key astrophysical features. A follow-up study is currently underway using more complex and realistic mock configurations, where results will also be compared to manual modeling outcomes.

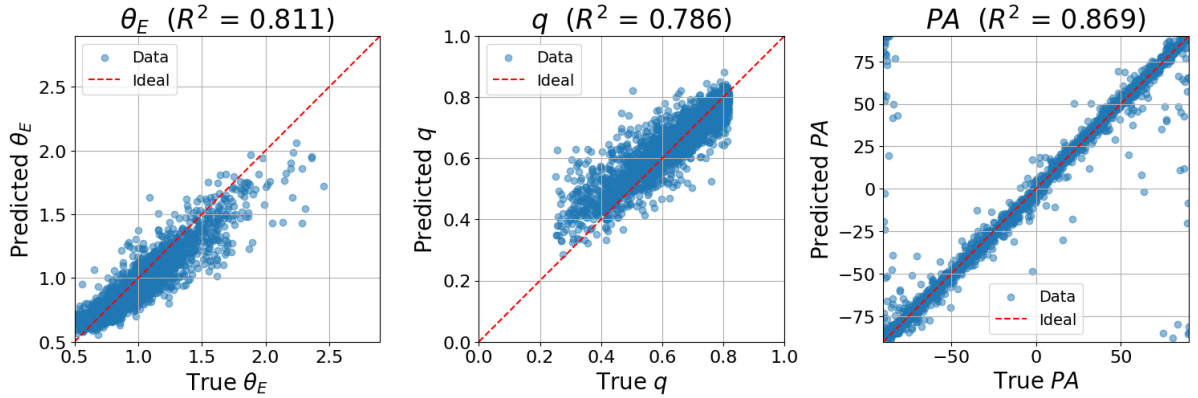

FIGURE 4. Comparison between predicted and true values for the three lensing parameters: Einstein radius (θ_E), axis ratio (q), and position angle (PA). The red dashed line denotes the ideal prediction.

CONCLUSION

In this study, we developed and evaluated a custom CNN to estimate strong gravitational lensing parameters—namely the Einstein radius, axis ratio, and position angle—directly from image data. Despite employing relatively simple lens models and assumptions for the lens galaxy, the network achieved high predictive accuracy for θ_E and q, demonstrating strong generalization performance on mock lensing images. Position angle predictions were less precise, likely due to intrinsic angular degeneracies near $\pm 90°$, a well-known challenge in lens modeling. Among the tested configurations, SiLU activation was more effective than PReLU in mitigating overfitting, and spatial dropout provided stronger regularization than standard dropout, particularly in deeper convolutional layers. Interestingly, using only three input channels (*gri*) outperformed the full five-band input (*grizy*), highlighting the importance of input dimensionality and filter selection. The Huber loss function proved more robust than MSE, particularly under skewed parameter distributions and outliers.

These results demonstrate the potential of CNNs as a reliable tool for automated gravitational lens modeling, especially when the network architecture, regularization strategy, and data representation are carefully optimized As a next step, the model will be applied to real lens candidates to assess its performance beyond simulations. In the long term, this modeling pipeline may be adapted for use in large-scale survei (e.g. LSST, DESI, etc.) systems to assist in the rapid analysis of gravitational lensing data.

ACKNOWLEDGMENTS

This work supported by the Program Penelitian, Pengabdian Masyarakat, dan Inovasi (PPMI) ITB 2025.

The Hyper Suprime-Cam (HSC) collaboration includes the astronomical communities of Japan and Taiwan, and Princeton University. The HSC instrumentation and software were developed by the National Astronomical Observatory of Japan (NAOJ), the Kavli Institute for the Physics and Mathematics of the Universe (Kavli IPMU), the University of Tokyo, the High Energy Accelerator Research Organization (KEK), the Academia Sinica Institute for Astronomy and Astrophysics in Taiwan (ASIAA), and Princeton University. Funding was contributed by the FIRST program from Japanese Cabinet Office, the Ministry of Education, Culture, Sports, Science and Technology (MEXT), the Japan Society for the Promotion of Science (JSPS), Japan Science and Technology Agency (JST), the Toray Science Foundation, NAOJ, Kavli IPMU, KEK, ASIAA, and Princeton University.

This paper makes use of software developed for the Large Synoptic Survey Telescope. We thank the LSST Project for making their code available as free software at http://dm.lsst.org

The Pan-STARRS1 Surveys (PS1) have been made possible through contributions of the Institute for Astronomy, the University of Hawaii, the Pan-STARRS Project Office, the Max-Planck Society and its participating institutes, the Max Planck Institute for Astronomy, Heidelberg and the Max Planck Institute for Extraterrestrial Physics, Garching, The Johns Hopkins University, Durham University, the University of Edinburgh, Queen's University Belfast, the Harvard-Smithsonian Center for Astrophysics, the Las Cumbres Observatory Global Telescope Network Incorporated, the National Central University of Taiwan, the Space Telescope Science Institute, the National Aeronautics and Space Administration under Grant No. NNX08AR22G issued through the Planetary Science Division of the NASA Science Mission Directorate, the National Science Foundation under Grant No. AST-1238877, the University of Maryland, and Eotvos Lorand University (ELTE) and the Los Alamos National Laboratory.

Based [in part] on data collected at the Subaru Telescope and retrieved from the HSC data archive system, which is operated by Subaru Telescope and Astronomy Data Center at National Astronomical Observatory of Japan.

REFERENCES

[1] J. Wagner, "A model-independent characterisation of strong gravitational lensing by observables," *Universe*, vol. 5, no. 177, 2019, doi: 10.3390/universe5070177.

[2] S. Schuldt *et al.*, "HOLISMOKES X: Comparison between neural network and semi-automated traditional modeling of strong lenses," *Astron. Astrophys.*, vol. 673, p. A33, 2023, doi: 10.1051/0004-6361/202244534.

[3] C. Schaefer, M. Geiger, T. Kuntzer, and J.-P. Kneib, "Deep convolutional neural networks as strong gravitational lens detectors," *Astron. Astrophys.*, vol. 611, p. A2, 2018, doi: 10.1051/0004-6361/201731201.

[4] A. Davies, S. Serjeant, and J. M. Bromley, "Using convolutional neural networks to identify gravitational lenses in astronomical images," *Mon. Not. R. Astron. Soc.*, vol. 487, no. 4, pp. 5263–5271, 2019, doi: 10.1093/mnras/stz1712.

[5] R. Pearce-Casey *et al*., "Euclid: Searches for strong gravitational lenses using convolutional neural networks in early release observations of the Perseus field," *Astron. Astrophys.*, vol. 696, p. A214, 2025, doi: 10.1051/0004-6361/202453152.
[6] Y. D. Hezaveh, L. P. Levasseur, and P. J. Marshall, "Fast automated analysis of strong gravitational lenses with convolutional neural networks," *Nature*, vol. 548, pp. 555–557, 2017, doi: 10.1038/nature23463.
[7] S. Schuldt *et al*., "HOLISMOKES IX: Neural network inference of strong-lens parameters and uncertainties from ground-based images," *Astron. Astrophys.*, vol. 671, p. A16, 2023, doi: 10.1051/0004-6361/202244325.
[8] S. Huber and S. H. Suyu, "HOLISMOKES XII: Time-delay measurements of strongly lensed Type Ia supernovae using a long short-term memory network," *Astron. Astrophys.*, vol. 692, p. A15, 2024, doi: 10.1051/0004-6361/202449952.
[9] S. Birrer and A. Amara, "lenstronomy: Multi-purpose gravitational lens modelling software package," *Phys. Dark Universe*, vol. 22, pp. 189–201, 2018, doi: 10.1016/j.dark.2018.11.002.
[10] K. C. Wong *et al*., "Survey of gravitationally lensed objects in HSC imaging (SuGOHI). VIII: New galaxy-scale lenses from the HSC SSP," *Publ. Astron. Soc. Jpn.*, vol. 74, pp. 1209–1219, 2022, doi: 10.1093/pasj/psac065.
[11] P. Gawade *et al*., "Neural network prediction of model parameters for strong lensing samples from the Hyper Suprime-Cam Survey," *Mon. Not. R. Astron. Soc.*, vol. 540, pp. 3384–3399, 2025, doi: 10.1093/mnras/staf935.
[12] R. C. Pinto and A. R. Tavares, "PReLU: Yet another single-layer solution to the XOR problem," *arXiv*, 2024, doi: 10.48550/arXiv.2409.10821.
[13] J. Tompson, R. Goroshin, A. Jain, Y. LeCun, and C. Bregler, "Efficient object localization using convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2015, pp. 648–656, doi: 10.1109/CVPR.2015.7298664
[14] K. Gokcesu and H. Gokcesu, "Generalized Huber loss for robust learning and its efficient minimization for a robust statistics," *arXiv*, 2021, doi: 10.48550/arXiv.2108.12627.
[15] H. Hoshino *et al*., "Luminous red galaxies in clusters: Central occupation, spatial distributions and miscentring," *Mon. Not. R. Astron. Soc.*, vol. 452, pp. 998–1013, 2015, doi: 10.1093/mnras/stv1308.
[16] A. T. Jaelani *et al*., "Survey of Gravitationally lensed Objects in HSC Imaging (SuGOHI) − X. Strong lens finding in the HSC-SSP using convolutional neural networks," *Mon. Not. R. Astron. Soc.*, vol. 535, pp. 1625–1639, 2024, doi: 10.1093/mnras/stae2442.
[17] M. Tanaka, "Photometric redshift with Bayesian priors on physical properties of galaxies," *Astrophys. J.*, vol. 801, no. 1, p. 20, 2015, doi: 10.1088/0004-637X/801/1/20.
[18] M. Meneghetti, *Introduction to Gravitational Lensing: With Python Examples*. Cham, Switzerland: Springer, 2021, doi: 10.1007/978-3-030-73582-1.
[19] R. Kormann, P. Schneider, and M. Bartelmann, "Isothermal elliptical gravitational lens models," *Astron. Astrophys.*, vol. 284, pp. 285–299, 1994.
[20] A. S. Bolton *et al*., "The Sloan Lens ACS Survey. V. The full ACS strong-lens sample," *Astrophys. J.*, vol. 682, no. 2, pp. 964–984, 2008, doi: 10.1086/589327.
[21] C. R. Keeton, C. S. Kochanek, and U. Seljak, "Shear and ellipticity in gravitational lenses," *Astrophys. J.*, vol. 482, no. 2, pp. 604–620, 1997, doi: 10.1086/304172.
[22] E. O. Ofek, H.-W. Rix, and D. Maoz, "The redshift distribution of gravitational lenses revisited: Constraints on galaxy mass evolution," *Mon. Not. R. Astron. Soc.*, vol. 343, no. 2, pp. 639–652, 2003, doi: 10.1046/j.1365-8711.2003.06707.x.
[23] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," *arXiv*, 2015, doi: 10.48550/arXiv.1511.08458.
[24] J. Wu, "Introduction to convolutional neural networks," Nanjing Univ., Tech. Rep., 2017.
[25] M. Lee, "GELU activation function in deep learning: A comprehensive mathematical analysis and performance," *arXiv*, 2023, doi: 10.48550/arXiv.2305.12073.
[26] C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, "Activation functions: Comparison of trends in practice and research for deep learning," *arXiv*, 2018, doi: 10.48550/arXiv.1811.03378.