Strong Gravitational Lensing in the Era of Big Data
Proceedings IAU Symposium No. 381, 2024
H. Stacey, A. Sonnenfeld & C. Grillo, eds.
doi:10.1017/S1743921323003708

Lens Discovery in the Era of Wide-area Surveys

Philip Holloway[1], **Philip J Marshall**[2,3] **and Aprajita Verma**[1]

[1]Sub-department of Astrophysics, University of Oxford, Denys Wilkinson Building, Keble Road, Oxford, OX1 3RH, UK. email: philip.holloway@physics.ox.ac.uk

[2]Kavli Institute for Particle Astrophysics and Cosmology, Department of Physics, Stanford University, Stanford, CA 94305, USA

[3]SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA

Abstract. Forthcoming data from the Vera Rubin Observatory, *Euclid* and *Roman* telescopes are expected to increase the number of strong lenses by two orders of magnitude. With current discovery methods these would be accompanied by an even greater number of false positives. In that context we find that using an ensemble of classifiers would provide a more complete sample of high-purity lenses and present methods to post-process the outputs of such classifiers to give reliable probabilities that a given image contains a lens.

Keywords. gravitational lensing: strong, galaxies: general, methods: data analysis

1. Introduction

With the arrival of wide-area telescopes such as the Vera Rubin Observatory, Euclid and Roman space telescopes, the number of strong lenses identified will increase to $\sim 10^5$ (Collett (2015), Holloway *et al.* (2023)). Current lens detection techniques require significant time-investment to remove false positives identified by automated or human classifiers which will only increase when such telescopes come online. This motivates our investigation into methods for improving the performance of strong lens finding.

2. Data and Method

To develop and test our method, we use outputs from two strong lens classifiers applied to Hyper-Suprime Cam (HSC) data. These classifiers are 1) a neural network (HOLISMOKES VI, Cañameras *et al.* (2021)) and 2) citizen science classifications from Space Warps (SuGOHI VI, Sonnenfeld *et al.* (2020)). These were cross-matched to within 1", producing $\sim 110,000$ galaxies for which outputs from both classifiers were available. For a subset of 3,514 typically high-scoring objects, grades were available following subsequent visual inspection. We used these grades as a 'ground-truth' to determine the performance of each classifier.

We mapped the classifier output to calibrated probability as follows. We took the distribution of grade A+B candidates (here considered true lenses) as a function of classifier output ranking and applied the following procedures to determine this mapping for each classifier: isotonic regression (Zadrozny & Elkan (2002)), variable bin fitting (akin to a moving average, but with a fixed number of lenses per bin), and the Kullback-Leibler Importance Estimation Procedure (KLIEP, Sugiyama *et al.* (2008)). These calibration mappings are validated against a separate validation set in Figures 1a & 1b which show

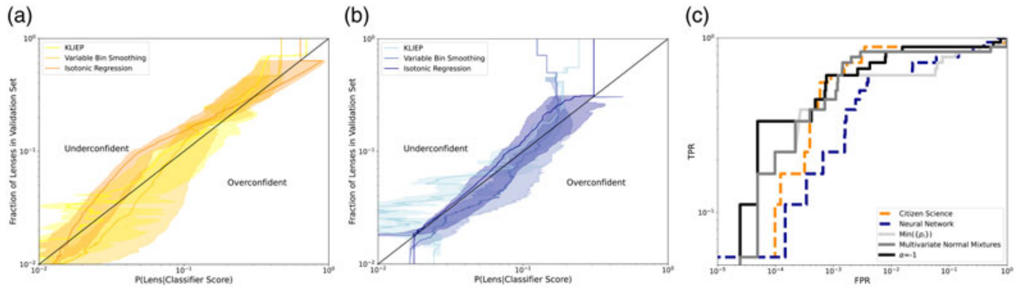

Figure 1. a & b) Validation for the 3 calibration methods applied to (a) the citizen science classifier and (b) the neural network: a perfectly calibrated classifier would lie along the $y = x$ line. c) Receiver operating characteristic (ROC) curve for the individual lens classifiers (dashed) and combined methods. Ungraded subjects were treated as non-lenses for this plot.

that across a wide region of probability-space, such a mapping can provide an accurate probability a given object is a lens.

To combine the calibrated outputs we used a generalised mean of the form: $P(p_i, \alpha) = (\frac{1}{N}\sum_i^N p_i^\alpha)^{\frac{1}{\alpha}}$ where p_i denotes the calibrated outputs from the N different classifiers, and α is a tuneable parameter; $\alpha \to \pm\infty$ corresponds to $Max(\{p_i, ...p_N\})$ and $Min(\{p_i, ...p_N\})$. We also trialled Bayesian probability combination using multivariate normal mixtures detailed in Pirš & Štrumbelj (2019). The results of the best performing methods are shown in Figure 1c. We find at the high-purity end of the curve (FPR$\sim 10^{-4}$), combining the two classifiers can roughly double the expected completeness from $\sim 15\%$ to $\sim 30\%$ though the ensemble classifier doesn't improve significantly that of the citizen science for larger false positive rates. With forthcoming wide-area surveys, this will help reduce the amount of time-consuming visual inspection required to verify lens candidates. A more in-depth analysis and further results will be presented in a forthcoming paper, Holloway *et al.* (2023, in prep).

3. Conclusion

We find the outputs of strong lens classifiers can be post-processed to give reliable probabilities that a given image contains a lens allowing the accurate ranking of lens candidates identified across multiple different methods. We find an ensemble classifier would provide a more complete sample of lenses for a given (high) value of purity than the individual classifiers by themselves. This will allow a larger sample of strong lenses to be used for future population-level analysis enabled by LSST, *Euclid* and *Roman* surveys.

4. Acknowledgements

With thanks to Raoul Cañameras and Alessandro Sonnenfeld for providing data for this project.

Supplementary material

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/S1743921323003708

References

Collett, T. E. 2015, *ApJ*, 811, 20
Holloway, P., Verma, A., Marshall, P. J., More, A., Tecza, M. 2023, *MNRAS*, 525, 2341–2354
Cañameras, R., Schuldt, S., Shu, Y., *et al.* 2021 *A&A*, 653, L6

Sonnenfeld, A., Verma, A., More, A., *et al.* 2020 *A&A*, 642, A148
Zadrozny, B., Elkan, C. 2002, Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 694–699
Sugiyama, M., Suzuki, T., Nakajima, S. *et al.* 2008, Annals of the Institute of Statistical Mathematics, 60, 699
Pirš, G., Štrumbelj, E. 2019, Journal of Machine Learning Research, 20, 1892–1909