Lens Discovery in the Era of Wide-area Surveys

Philip Holloway[1,2], Philip J Marshall[2], Aprajita Verma[1]

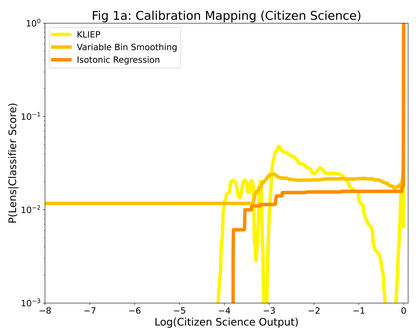
1. University of Oxford, UK
2. Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, USA

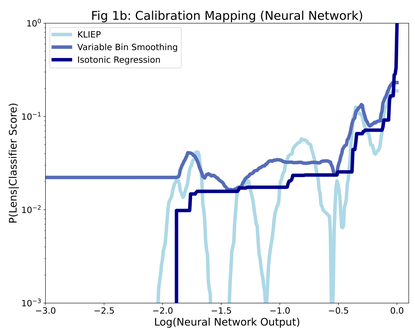
Summary

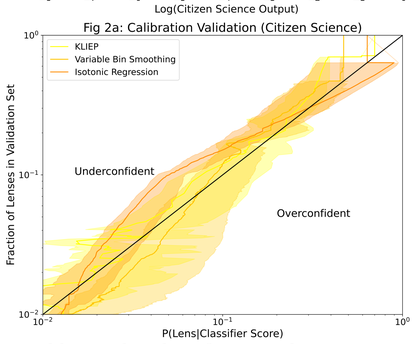
With the arrival of wide-area telescopes such as the Vera Rubin Observatory, Euclid and Roman space telescopes, the number of strong lenses identified will increase to ~10^5 (Collett 2015, Holloway 2023 in prep). Current lens detection techniques require significant time-investment to remove false positives identified by automated or human classifiers. Our work aims to do the following:

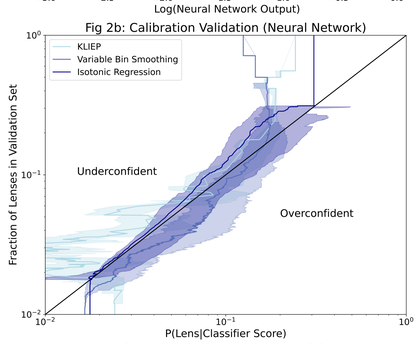
1. Produce calibrated probabilities that a given system is strongly lensed.
2. Identify and test a methodology to combine multiple strong lens classifiers, aiming to maximise purity without significant compromises on completeness.

We can calibrate strong-lens classifiers to produce accurate probabilities a given object is a lens



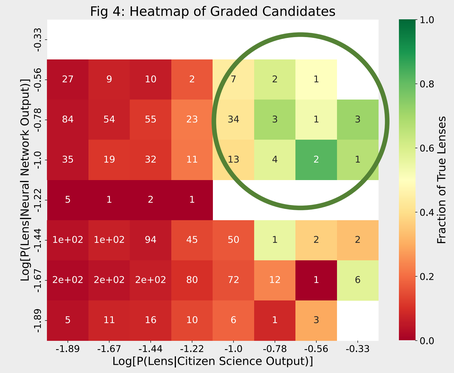


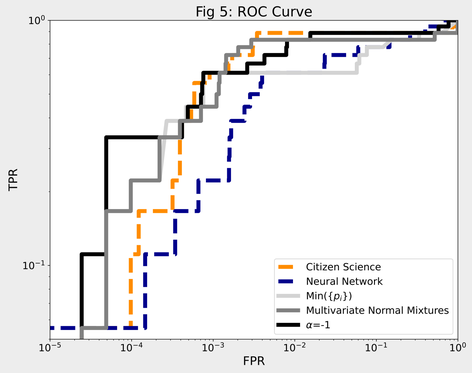




We calibrate the classifiers via the mappings above (top row), then validate this calibration on a separate validation set (bottom row). A perfectly calibrated classifier would lie along the y=x line.

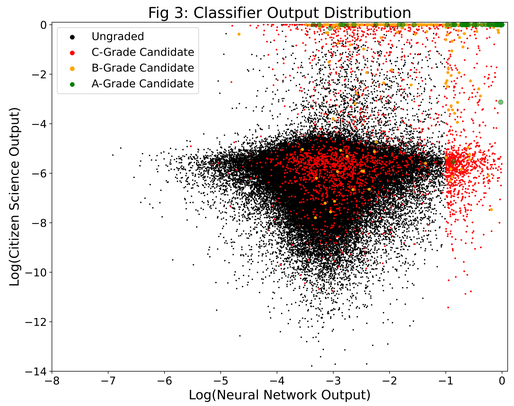
Higher purity can be achieved through combining multiple strong lens classifiers





The left hand plot shows the distribution of classified objects as a function of their calibrated citizen science and neural network scores, shaded by the fraction of true lenses in each bin. The annotations indicate the number of graded objects in each bin. The purity of lenses is greatest where both classifiers give high scores. This is reflected in the ROC curve on the right, for which combining the scores provides the highest TPR in the low FPR (i.e. high purity) regime. The harmonic mean ($\alpha = -1$) is the best performing when prioritising purity, however most of the combined methods show improved purity over the individual classifiers.

What do we gain from combining classifiers?



There are a significant number of objects for which the two classifiers disagree, suggesting there is scope for an ensemble classifier which can find the 'best of both worlds'.

What data are we using?

To develop and test our method, we use outputs from two strong lens classifiers on Hyper-Suprime Cam (HSC) data. These classifiers are:

- A neural network (HOLISMOKES VI, Cañameras et al. (2021)) applied to 5.4×10^7 galaxies in HSC PDR2.
- Citizen science classifications from Space Warps (Sonnenfeld et al. (2020)) of ~300,000 objects using HSC S17A.

These were cross-matched to within 1", producing ~110,000 galaxies for which both classifier outputs 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. Figure 3 shows the distribution of scores for cross-matched objects.

What's our method?

We first produced a mapping from classifier output to calibrated probability. We took the distribution of grade A+B candidates (considered true lenses) as a function of classifier output ranking (hereafter 'grade distribution') and applied the following procedures to determine this mapping for each classifier, show in Figure 1: *Isotonic regression* (fitting a monotonically increasing curve to the grade distribution), *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 2008, a form of Gaussian mixture model). These calibration mappings are validated in Figure 2.

We combined the calibrated outputs to maximise the purity of the resultant sample. We first used a generalised mean of the form: $P(p_i, \alpha) = \left(\frac{1}{N}\sum_{i=1}^{N} p_i^{\alpha}\right)^{1/\alpha}$ where p_i denotes the calibrated outputs from the N different classifiers, and α is a tuneable parameter. $\alpha = 1$ corresponds to the usual geometric mean, while $\alpha \to \pm\infty$ correspond to $Max(\{p_i, \dots, p_N\})$ and $Min(\{p_i, \dots, p_N\})$. 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 5.

Acknowledgements & References

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

Collett T., **2015**, ApJ, 811, 20
Cañameras R., et al., **2021**, A&A, 653, L6
Sonnenfeld A., et al., **2020**, A&A, 642, A148
Sugiyama M., Suzuki T., Nakajima S., Kashima H., Bünau P. V., Kawanabe M., **2008**, Annals of the Institute of Statistical Mathematics, 60, 699
Pirš G., Štrumbelj E., 2019, J. Mach. Learn. Res., 20, 1892–1909