Unveiling the most dusty and distant galaxies in wide-field infrared imaging with machine learning

Supervision team: Stephen Serjeant, Hugh Dickinson, Jane Bromley

Lead supervisor: Stephen Serjeant, stephen.serjeant@open.ac.uk

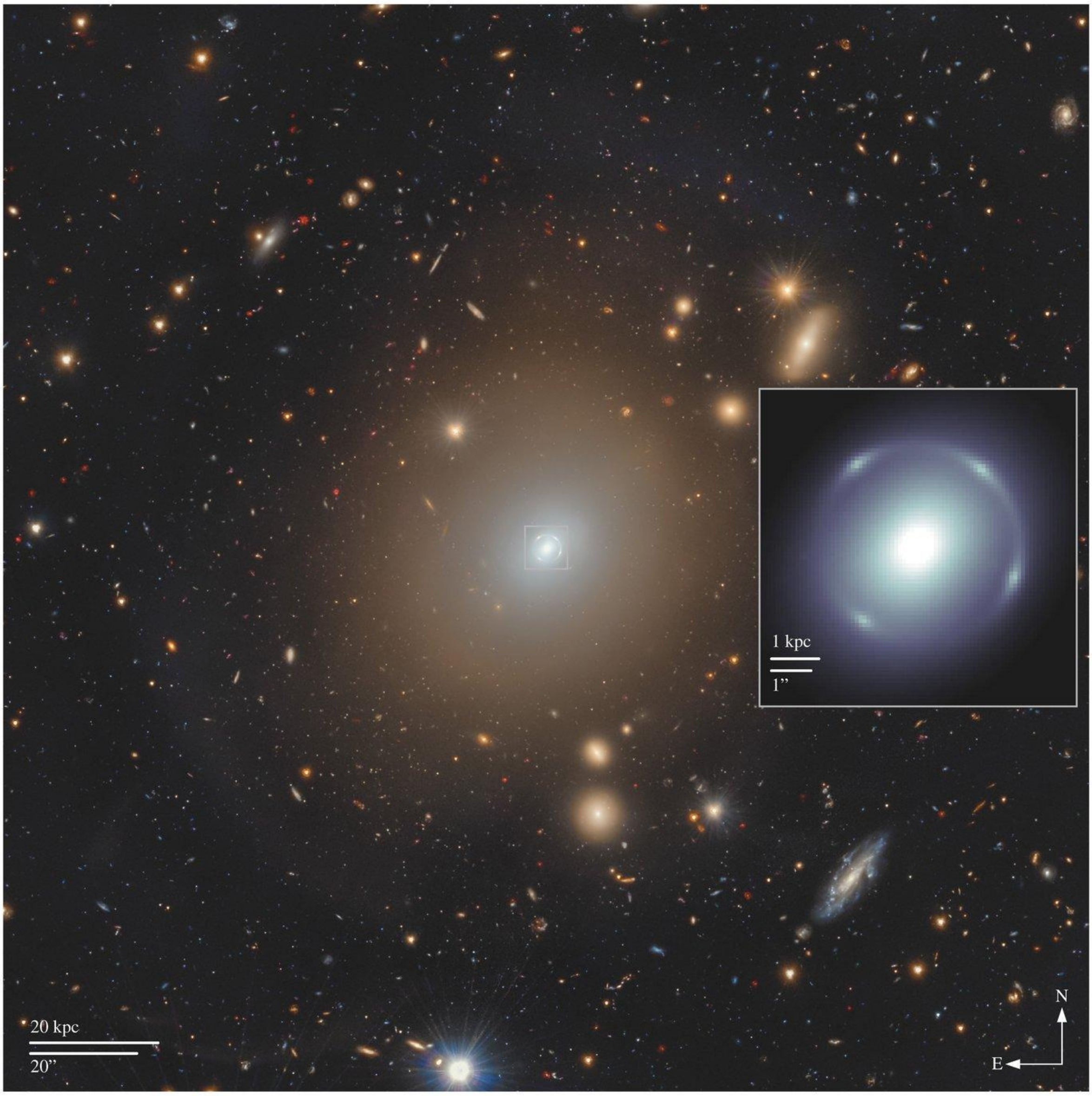

Image credit: ESA/Euclid/Euclid Consortium/NASA, image processing by Tian Li

Project highlights:

- Project importance: to find large, pure samples of rare strong gravitational lenses in current and forthcoming giant astronomical sky surveys, including and especially the pathological examples that are under-represented to date in near-infrared surveys, such as ultra-high-redshift galaxies that reionised the Universe, and dusty star-forming galaxies that account for half the cosmic budget of star formation. The high angular resolution afforded by gravitational lensing, ~2-50x the unlensed case, will provide a window on the structures of these galaxies and the physical processes that drive their evolution.

- Outcomes: to discover rare strongly gravitationally lensed galaxies in the Euclid imaging surveys
- Career development, partners and collaborators: the project will involve joining working groups of the Euclid space telescope and the LSST project on the Vera Rubin Observatory, which will give the student visibility in large international research consortia to promote their professional development.

Project description:

The Euclid space telescope is mapping about a third of the sky to nearly Hubble Space Telescope (HST) quality. Euclid has begun an avalanche of discoveries of rare gravitational lenses, increasing the number of known lenses by over a factor of one hundred, illuminating both the structure of foreground dark matter halos and the structures and properties of the distant galaxies. The huge increase is especially useful for rare lensing events, such as the brightest red lensed galaxies that are prime candidates for JWST/ALMA follow-ups, and which will be mostly from the dusty star forming galaxy population around Cosmic Noon where the cosmic star formation history peaked.

There are many directions that this PhD project could take, because of the rich abundance of data and our successful track record with developing relevant machine learning tools; we outline a few here.

Euclid's ~100,000 strong gravitational lenses have to be found in a catalogue of about a billion unlensed galaxies. We have solved most of this 'needle in a haystack' problem using a convolutional neural net trained on 100 million volunteer classifications (Pearce-Casey et al. 2025). The resulting catalogues are still only about 10 percent pure, meaning that around a million candidates would need careful visual inspection; this is possible, but more progress could be made by incorporating a filtering stage of automated lens modelling, or using some other technique to enforce the symmetries of the underlying physics. One direction for this PhD project is to improve the purity of the machine learning classifications by making better use of the geometry and astrophysics of gravitational lensing, and of galaxies in general.

A major advantage of having a very large new catalogue of gravitational lenses is that it is an excellent way in principle to find rare types of object, but this also poses a problem for machine learning if it has not been trained to find the rare systems. The Euclid and HST imaging of known lensed dusty galaxies would often fail an expert inspection looking for lensing, if that expert did not already know it was a lens. Euclid's wide survey, cross-correlated against infrared surveys with e.g. Herschel and the JCMT, is an excellent starting point for creating imaging catalogues of dusty star-forming galaxies that can be used for lensing simulations for machine learning training, which is another possible initial direction for this PhD project. Again, being able to start with a working lens detection pipeline and fine-tune it for the problem in hand is a significant time-saver.

Finally, a significant advantage of strong lensing is the angular resolution improvements that provide an important window on the distant galaxy population, but not all the observations have the same quality. For example, the Euclid NISP near-infrared instrument has an angular resolution around a factor of two worse than Euclid's visible VIS instrument. Nevertheless, it may be that astrophysically-important information is at the threshold of being detected in the near-infrared, such as the existence of dusty star-forming clumps in the background galaxies that yield insights on the physical processes that drive the dramatic evolution in the cosmic star formation history. Machine learning has many tools that can assist, and even if individual reconstructions are not completely reliable, they may still usefully determine the population properties when combined with simulations to measure the recovered clump reliability and completeness (i.e. purity and recall). There is a great deal of successful work on which to build, if you would like to prioritise the super-resolution and image deconvolution aspects of this project. We have previously shown that far-infrared and submillimetre-wave imaging from the ESA Herschel space observatory can be very effectively deconvolved using a denoising autoencoder machine learning architecture (Lauritsen et al. 2021), and we've deployed the same technology on simulated data for the proposed NASA PRIMA space telescope (Donnellan et al. 2024). We also currently have postdoctoral researchers working on super-resolution of Rubin LSST gravitational lensing imaging, and of Euclid near-infrared imaging of galaxies in general, using the Pix2pix machine learning architecture.

References:

1. Donnellan et al. 2024 https://arxiv.org/abs/2404.06935
2. Lauritsen et al. 2021 https://ui.adsabs.harvard.edu/abs/2021MNRAS.507.1546L
3. Pearce-Casey et al. 2025 https://ui.adsabs.harvard.edu/abs/2025A%26A...696A.214P

Qualifications required: Minimum BSc 2:1 in a relevant discipline; recommended MPhys 2:1 or above.