A&A, 698, A264 (2025)
https://doi.org/10.1051/0004-6361/202453195
© The Authors 2025

Astronomy & Astrophysics

HOLISMOKES

XV. Search for strong gravitational lenses combining ground-based and space-based imaging

A. Melo[1,2,⋆], R. Cañameras[3,1,2], S. Schuldt[4,5], S. H. Suyu[2,1], I. T. Andika[2,1], S. Bag[2,1], and S. Taubenberger[2,1]

[1] Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, 85748 Garching, Germany
[2] Technical University of Munich, TUM School of Natural Sciences, Physics Department, James-Franck-Straße 1, 85748 Garching, Germany
[3] Aix Marseille Univ, CNRS, CNES, LAM, Marseille, France
[4] Dipartimento di Fisica, Universita degli Studi di Milano, Via Celoria 16, I-20133 Milano, Italy
[5] INAF – IASF Milano, Via A. Corti 12, I-20133 Milano, Italy

Received 27 November 2024 / Accepted 10 April 2025

ABSTRACT

In the past, researchers have mostly relied on single-resolution images from individual telescopes to detect gravitational lenses. We present a search for galaxy-scale lenses that, for the first time, combines high-resolution single-band images (in our case from the Hubble Space Telescope, HST) with lower-resolution multiband images (in our case from the Legacy survey, LS) using machine learning. This methodology simulates the operational strategies employed by future missions, such as combining the images of Euclid and the Rubin Observatory's Legacy Survey of Space and Time (LSST). To compensate for the scarcity of lensed galaxy images for network training, we generated mock lenses by superimposing arc features onto HST images, saved the lens parameters, and replicated the lens system in the LS images. We tested four architectures based on ResNet-18: (1) using single-band HST images, (2) using three bands of LS images, (3) stacking these images after interpolating the LS images to HST pixel scale for simultaneous processing, and (4) merging a ResNet branch of HST with a ResNet branch of LS before the fully connected layer. We compared these architecture performances by creating receiver operating characteristic (ROC) curves for each model and comparing their output scores. At a false-positive rate of 10^{-4}, the true-positive rate is ∼0.41, ∼0.45, ∼0.51 and ∼0.55, for HST, LS, stacked images and merged branches, respectively. Our results demonstrate that models integrating images from both the HST and LS significantly enhance the detection of galaxy-scale lenses compared to models relying on data from a single instrument. These results show the potential benefits of using both Euclid and LSST images, as wide-field imaging surveys are expected to discover approximately 100 000 lenses.

Key words. gravitational lensing: strong – methods: data analysis

1. Introduction

Gravitational lensing occurs when a massive object, such as a galaxy or cluster of galaxies, bends and magnifies the light from a more distant object, such as a quasar or another galaxy. This effect can produce multiple images, arcs, or even rings of the background object in the case of strong lensing, providing a powerful tool for studying the distribution of dark matter and understanding the nature of dark energy (Schneider et al. 1992). Moreover, lensed time-variable objects like quasars or supernovae (SNe) can be used to measure the Hubble constant (H_0) and cosmological parameters, as proposed by Refsdal (1964). Our HOLISMOKES program, which stands for Highly Optimized Lensing Investigations of Supernovae, Microlensing Objects, and Kinematics of Ellipticals and Spirals (Suyu et al. 2020), aims to discover lensed SNe, determine the expansion rate of the Universe and study the progenitor systems of the SNe. The latter is possible by capturing the SN image at very early times (Chen et al. 2022). While only a few lensed SNe have been identified so far (Kelly et al. 2015, 2022; Goobar et al. 2017, 2022; Polletta et al. 2023; Frye et al. 2024; Rodney et al. 2021; Chen et al. 2022; Pierel et al. 2024), the number is expected to increase significantly with the wide-field imaging survey by the Rubin Observatory Legacy Survey of Space and Time (LSST, Ivezic et al. 2008) and the Euclid space telescope (Laureijs et al. 2011). Forecasts for these upcoming surveys suggest a substantial increase by a factor of several hundreds in the discovery of lensed SNe (Oguri & Marshall 2010; Wojtak et al. 2019; Goldstein et al. 2019; Arendse et al. 2024; Bag et al. 2024; Sainz de Murieta et al. 2024).

Over the past few decades, significant efforts have been made to detect and study gravitational lenses. Deep, wide-scale surveys have played a crucial role in finding new gravitational lenses, such as the Sloan Digital Sky Survey (SDSS, Bolton et al. 2006; Oguri et al. 2006; Brownstein et al. 2012), the Cosmic Lens All-Sky Survey (CLASS; e.g., Myers et al. 2003), and the Survey of Gravitationally-lensed Objects in HSC Imaging (SuGOHI, Sonnenfeld et al. 2018; Chan et al. 2020; Jaelani et al. 2020; Wong et al. 2022; Jaelani et al. 2024; Chan et al. 2024). The development of automated algorithms and machine learning techniques has revolutionized the field, allowing for vast amounts of astronomical data to be processed and identifying lens candidates with great efficiency and accuracy (e.g., Lanusse et al. 2018; Jacobs et al. 2019). In particular,

⋆ Corresponding author: amelo@mpa-garching.mpg.de

Open Access article, published by EDP Sciences, under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
This article is published in open access under the Subscribe to Open model. Open access funding provided by Max Planck Society.

Convolutional Neural Networks (CNNs) have become increasingly important in astronomy due to their effectiveness in processing multi-wavelength imaging data (e.g., Jacobs et al. 2019; Metcalf et al. 2019; Petrillo et al. 2019; Cañameras et al. 2020; Rojas et al. 2022; Zaborowski et al. 2023; Schuldt et al. 2025; Storfer et al. 2024).

The Hubble Space Telescope (HST) has been crucial in providing high-resolution images for detailed studies of known gravitational lenses and confirming new ones. Its capabilities have enabled researchers to discern intricate details in lens systems, resulting in improved constraints on lens models and the properties of the lensed sources (see, e.g., the review by Treu 2010; Shajib et al. 2024). Faure et al. (2008) identified lens candidates in the COSMOS field (Capak et al. 2007; Scoville et al. 2007), while Pourrahmani et al. (2018) expanded on this by employing a CNN to search for lenses on single-HST/ACS i-band observations of the COSMOS field. However, their resulting sample of strong-lens candidates suffered from high contamination, and the lack of color information made the visual inspection challenging. Single-band lens finding has limitations even at high resolutions, arguing for the inclusion of color information, even from lower-resolution ground-based images. Recent advancements have involved the use of multiband imaging from ground-based surveys, such as the Dark Energy Survey (DES), the Legacy Survey (LS), PanSTARRS, and the Hyper Suprime-Cam (HSC) (e.g., Diehl et al. 2017; Cañameras et al. 2020; Huang et al. 2020; Cañameras et al. 2021; Rojas et al. 2022). These surveys have helped identify new lens candidates through color and morphological analysis. However, all the images are from one single instrument, and none have combined images of different resolutions and in different bands to search for strong lenses in an automated way.

Looking ahead, Euclid's wide-field space telescope is conducting an extensive survey, expected to cover approximately 14 000 square degrees of the sky (Euclid Collaboration: Mellier et al. 2025). Being the first telescope to provide high-resolution images, even though just in a single optical band, for a large sky area, the large Euclid data set of galaxy images is suitable for gravitational lens searches (Laureijs et al. 2011). The searches of galaxy-scale lenses have already begun with the Early Release Observations data (Acevedo Barroso et al. 2025; Pearce-Casey et al., in prep.; Chowdhary et al., in prep.). Furthermore, high-resolution images show great promise, particularly in the discovery of a wider range of lens configurations (e.g., Garvin et al. 2022; Wilde et al. 2022). Implementing advanced CNN architectures into the Euclid pipeline will be essential for efficiently identifying new lenses and maximizing the scientific output of the mission (Petrillo et al. 2019). Even though these high-resolution images are essential for resolving detailed lensing features, such as multiple images and faint arcs, multiband imaging will increase the efficiency and reduce false positives (Metcalf et al. 2019).

Rubin LSST complements Euclid's high-resolution imaging by providing deep, multiband observations across six optical bands: u, g, r, i, z, and y (Ivezić et al. 2019). The LSST's ability to capture the sky repeatedly over its 10-year survey will create an unprecedented time-domain data set essential for identifying transient events and monitoring changes in known objects. By combining LSST's multiband data with Euclid's high-resolution images, the efficiency of gravitational lens searches should be significantly enhanced. This synergy between Euclid and LSST creates a powerful, complementary tool for the robust detection and confirmation of galaxy-scale lenses.

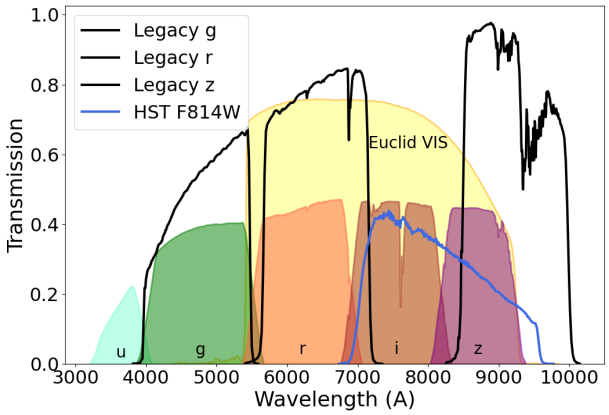

Fig. 1. Some of the filters from LSST (in shaded color labeled with u, g, r, i, and z) and Euclid (shaded yellow). The Legacy Survey g, r, and z filters, as well as the HST F814W filter, are shown in black and blue lines, respectively. Image based on Capak et al. (2019).

In this study, we present a CNN architecture based on ResNet-18 (He et al. 2016), trained on simulated and real lensing data, to improve the accuracy and efficiency of gravitational lens searches. While training solely on simulated data can lead to issues such as domain mismatch and overfitting to idealized conditions (Lanusse et al. 2018; Schaefer et al. 2018), we aim to overcome this by building a training set that incorporates real galaxy images. By combining high-resolution HST data with multiband LS images for the first time, our approach aims to improve detection capabilities, reduce false positives, and facilitate future applications in the Euclid and LSST missions. A similar multi-resolution strategy was explored by Bom et al. (2022), who used a dual-branch neural network to process simulated Euclid-like images with different resolutions, demonstrating that this approach can enhance lens detection. While their work focused on simulated data from a single instrument within a specific challenge setting, our study expands on this idea by using real training data and incorporating multiple instruments.

This paper is organized as follows. In Sect. 2, we introduce the construction of the ground truth data sets for training our CNNs, which includes simulating galaxy-scale lenses (for HST and LS). The network architectures and training are presented in Sect. 3. The results of applying the networks are presented in Sect. 4, along with a discussion about the architecture's performance, specially when combining images of different resolution. Finally, our main conclusions are in Sect. 5.

2. Data preparation

We simulated a sample of galaxy-scale lenses with two different resolutions, corresponding to a scenario with single-band space-based imaging and multiband ground-based imaging. Since Euclid provides high-resolution imaging primarily through its VIS channel (0.1″), which is most suitable for lens finding due to its superior resolution compared to the NISP channels (0.3″) (Collett 2015), we used the HST filter F814W image that has a pixel scale of 0.05″. For low-resolution multiband imaging, we used images from the DECam Legacy Survey (DECaLS) in the filters g, r, and z, which have a pixel scale of 0.263 arcseconds. These filters are selected to facilitate comparison with filters from the LSST and Euclid survey (see Fig. 1 for comparison), and are driven by the options currently available.

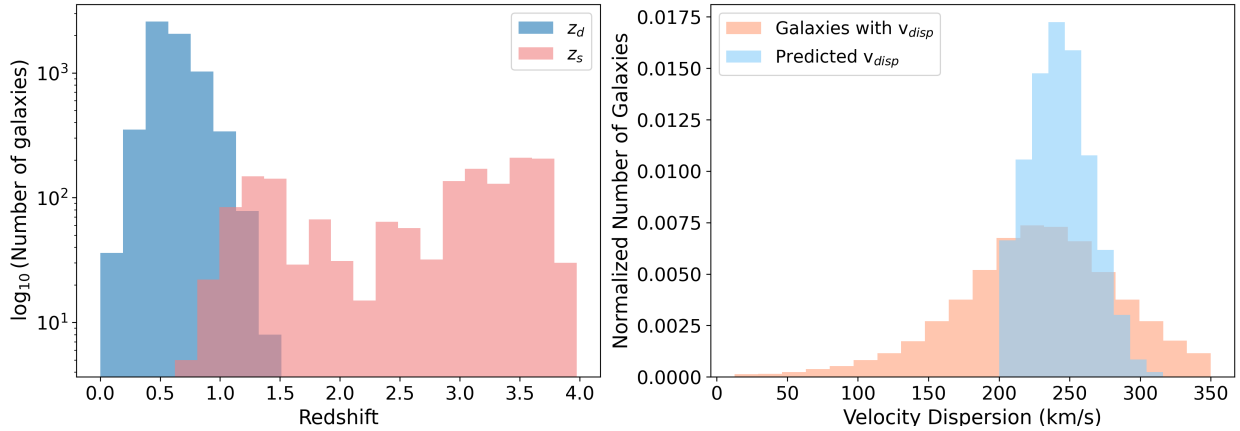

Fig. 2. Galaxy parameters (z_d, z_s, and v_{disp}) of the lenses and sources used for simulating galaxy-scale lenses. Left: histogram of the number of galaxies as a function of the redshift of the source z_s (red) and lens z_d (blue). Right: normalized v_{disp} distribution measured from SDSS used for training the KNN network (orange color), and the predicted normalized v_{disp} distribution for lens galaxies over the final sample of mocks (blue color).

Galaxy-scale lenses need to be simulated due to insufficient data for network training (deep learning typically requires more than 10 000 positive images for training). The process involved creating realistic arcs around real galaxies from LS and HST images, which act as lenses. It required several steps: selecting galaxies that act as the lenses, choosing the source images, and performing a strong lensing simulation. To ensure consistency between high-resolution and low-resolution mock images, we used the same lens parameters and source positions, creating identical lens configurations across resolutions.

First, we searched for luminous red galaxies (LRGs) to use as the lenses. These galaxies are preferred due to their high lensing cross-section (e.g., Turner et al. 1984) and because their smooth and uniform light profiles make it easier to separate the background emission from the lensed sources. We identified LRGs in the LS using the selection criteria from Zhou et al. (2023), accessed through the Astro Data Lab's query interface[1] (Nikutta et al. 2020). The color cuts in the selection criteria are based on the optical *grz* photometry from the Data Release 9 (DR9) and the WISE (Wright et al. 2010) *W*1 band in the infrared. After obtaining the list of LRGs, we crossmatched their coordinates with the HST Legacy Archive[2]; in cases of successful matches, we keep objects whose corresponding HST images have exposure times (total per object) of at least 100 seconds. The relatively low time limit is required to obtain an HST F814W image for at least ∼1/3 of the LS LRGs. With this criterion, we identify a total of approximately ∼10 000 LRG images in LS and HST. We then downloaded the 20″ × 20″ cutouts of the image.

For the background sources, we referred to the catalog created by Schuldt et al. (2021) of images of high-redshift galaxies from the Hubble Ultra Deep Field (HUDF, Inami et al. 2017). These galaxies are chosen due to their high spatial resolution (pixel size of 0.03″), high signal-to-noise ratio (SNR), availability of spectroscopic redshifts (Beckwith et al. 2006; Inami et al. 2017), and compatibility with the LS filters *g*, *r*, and *z*, corresponding to the HST filters F435W (λ = 4343.4Å), F606W (λ = 6000.8Å), and F850LP (λ = 9194.4Å), respectively. The spectroscopic redshifts of the selected HUDF galaxies falls in the range $0.1 < z < 4.0$. These images were used for simulations for each filter in the LS images, with color corrections applied to account for differences in filter transmission curves and photometric zero points. Specifically, source fluxes are rescaled using the difference in photometric zero points between HUDF and LS filters: $10^{0.4(zp_{LS}-zp_{HUDF})}$. Although we do not model the full SED of each galaxy, this correction provides a good approximation for simulations. Additionally, the HUDF *F*775*W* filter is used for the simulation in HST images, with color corrections to account for differences between filters.

We followed the procedure described in Schuldt et al. (2021) to simulate realistic galaxy-scale lenses by painting lens arcs first on the HST images. As input parameters, we required the redshift of the source (z_s), the redshift of the lens (z_d), and the velocity dispersion of the deflector (v_{disp}). Since not all LRGs have v_{disp} and redshift measured from the SDSS, we use the photometry and photometric redshift from LS to predict the velocity dispersion using a K-nearest neighbor (KNN) algorithm, following Rojas et al. (2022). We trained this algorithm with velocity dispersions using ∼580 000 LRGs with redshifts from SDSS and the redshifts and photometry from LS from our original list of LS without corresponding HST image. To keep the LRGs with reliable v_{disp}, we remove LRGs with $v_{disp} \leqslant 350\,\mathrm{km\,s^{-1}}$ and $v_{disp_error} \leqslant 50\,\mathrm{km\,s^{-1}}$ before training. As a result, we obtain a root-mean-square (rms) scatter in the prediction of $\sigma = 65\,\mathrm{km\,s^{-1}}$. The distribution of the z_d, z_s and v_{disp} from the input lens and source catalogs are shown in Fig. 2.

To create galaxy-scale lens simulations, we adopted a lens mass distribution with a Singular Isothermal Ellipsoid (SIE; Kassiola & Kovner 1993) profile using the input parameters: z_d, z_s, and v_{disp}. The lens centroid, axis ratio, and position angle were derived from the first and second brightness moments in the HST F814W images. The sources are randomly positioned in the source plane, and only simulations with an Einstein radius (θ_E) between 0.5″ and 2.5″ are considered (see Fig. 3). The choice of the range in θ_E took advantage of the high resolution provided by HST to study whether the architecture can effectively identify lensed systems with small θ_E, which were previously challenging to differentiate from non-lenses when using single-resolution images. Subsequently, the sources are lensed in the image plane using the GLEE software (Suyu & Halkola 2010; Suyu et al. 2012). Finally, for the HST mock simulation, the simulated lensed source image is convolved with the HST point spread function (PSF) model created using the Tiny Tim

[1] https://datalab.noirlab.edu/
[2] https://hla.stsci.edu/hlaview.html

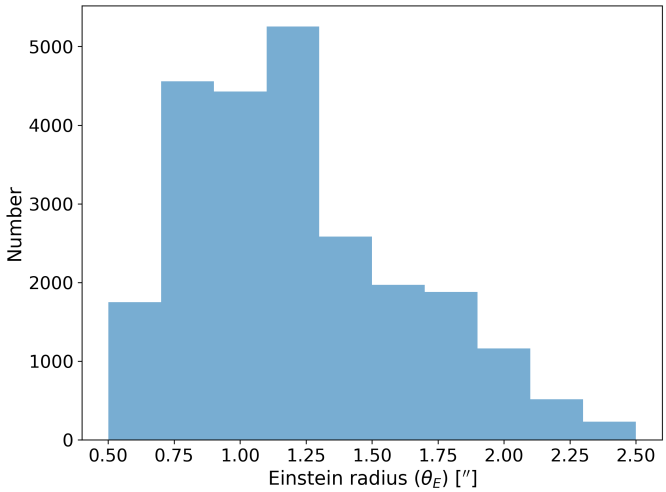

Fig. 3. Distribution of the Einstein radii of the mock strong lens systems used to train our networks.

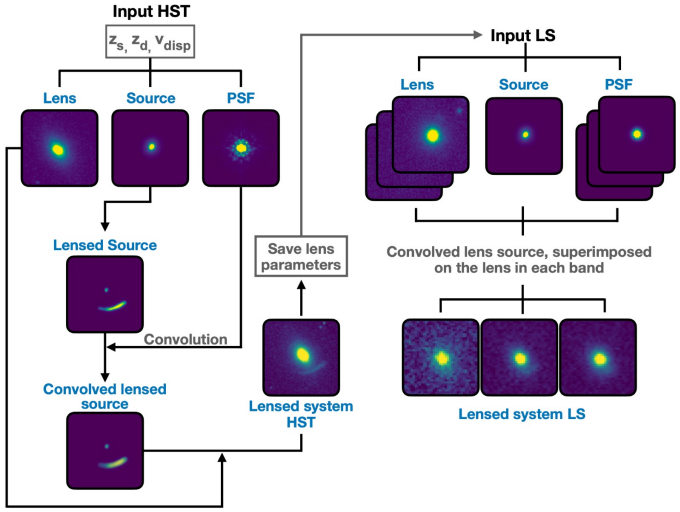

Fig. 4. Simulation pipeline procedure used for simulating galaxy-scale lenses. Further details are provided in Sec. 2.

package (Krist 1995), then rescaled using the HST zero-points, and added to the original LRG image to produce the final mock simulation. We refer to Schuldt et al. (2021) for additional details on the procedure. To simulate the same lens for ground-based multiband imaging of LS, we stored both the lens parameters and the fixed source position from the HST images, and used them to create the same lens configuration using the GLEE software for the LS g, r, and z images. We made sure to resample and rescale the HST source images using the LS PSF (computed using PSFEx and available through the Legacy Survey Viewer[3]). The simulation process is summarized in Fig. 4.

We initially obtain a total of 10 129 simulated lenses. However, upon comparing the simulated HST with the simulated LS images, we observe that not all arcs are visible in the LS images when using the same lens parameters. This outcome was expected due to the use of images with different resolutions and colors. To ensure the consistency in our sample, we selected simulations where the arcs are visible in all three LS bands and are at least 2σ above the sky background in the HST images (see

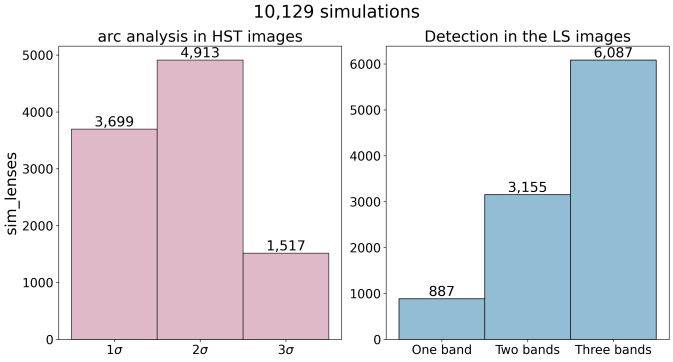

Fig. 5. Properties of the 10,129 HST and LS lens simulations: *Left*: number of mock lenses whose integrated arc brightness is 1σ, 2σ and 3σ above the sky background in HST images, and *Right*: number of mock lenses whose arcs are detected in LS images above 2σ. There are ~6000 mocks whose arcs are detected at $>2\sigma$ in HST and in all three LS bands.

Fig. 5 for further clarification of the whole sample). This left us with a total of ~6000 simulations.

Since this number is insufficient for training a network, we employed image rotation to the original LRG images (90, 180, and 270 degrees) and simulated new galaxy-scale lenses using different source images in each case. This created additional positive examples and effectively increased the data set size to approximately ~24 000 positive images. This process not only increased the size of the data set but also enhanced its variability, helping the model to become more robust to variations in orientation and configuration, which improves its generalization capabilities when applied to real astronomical data. The Einstein radius of the entire ~24 000 mock systems with HST and LS images is shown in Fig. 3; the overall distribution peaks at θ_E between 0.8 and 1.2″.

We used a diverse set of non-lens contaminants for the negative examples to help the networks learn the diversity of non-lens galaxies, with a particular focus on morphological types that resemble strong lenses (see Rojas et al. 2022; Cañameras et al. 2024). This includes various categories such as LRGs, ring galaxies, spiral galaxies, edge-on galaxies, groups, and mergers. The LRGs were obtained from the same parent sample that provides the lenses. Coordinates for the spiral and elliptical galaxies were obtained from the Galaxy Zoo[4] and then were used to download the images in HST and LS. In addition, galaxy classifications from the Galaxy Zoo: Hubble (GZH) project (Willett et al. 2017) and Galaxy Zoo DECaLS (Walmsley et al. 2022) are used to identify and download rings and mergers. All these negatives were then split into the training, validation, and test sets.

Finally, we split the positive data sample into 56% training, 14% validation, and 30% test sets (see Table 1). For the test set, we included ~8000 positive examples, which consisted of simulated galaxy-scale lenses, along with 364 real lens candidates obtained from the literature (Bolton et al. 2008; Auger et al. 2009; Pawase et al. 2014) that had both HST and LS images. Since we needed a larger sample of negative examples for the test set in order to obtain robust estimates of low false-positive rates (FPRs), especially in HST, we also used images of objects observed in the field-of-view (FOV) of the LRGs acting as the lenses. These contaminants were initially identified with a flux threshold above the background. Subsequently, the image was

[3] https://www.legacysurvey.org/viewer

[4] https://data.galaxyzoo.org/

Table 1.

HST/Legacy Survey	Training	Validation	Test set
Positive	~14 300	~3600	~8000
Negative	~20 000	~4300	~120 000

Notes. Distribution of samples in the HST/Legacy Survey data set for training, validation, and test sets.

cropped to $10'' \times 10''$ in both HST and LS. Later, we filtered the images to ensure that there were no zero values in the field. While we acknowledge that using LRGs from the same simulations as negative examples and extracting cutouts from the same FOV is not ideal-leading to potential repetitions in training and test sets-this approach is driven by the limited data available to us. Despite these limitations, the negatives provide a diverse and representative sample for training. The ~120 000 in the test set were a combination from both HST and LS data. Examples of positive images can be seen in Fig. 6. The final image used for training is $10'' \times 10''$, resulting in 200×200 pixels for HST images and 38×38 pixels for LS images.

3. Network training

We needed our network to be both effective and simple to accommodate training for four scenarios (see Fig. 7):

1. LS images (three images as input).
2. HST images (single image as input).
3. Stacked images (four images as input).
4. Merged branches (four images as input, but split into two branches).

To accomplish the research objective, we used ResNet-18, a deep CNN architecture introduced by He et al. (2016), in the four scenarios. This network has previously been used in lens search, for instance by Cañameras et al. (2021), who employed a network inspired by ResNet-18 to search for lenses in the HSC Survey, and also by Shu et al. (2022) (based on the work of Lanusse et al. 2018), who used it to search for high-redshift lenses in the HSC Survey. The architectures are slightly modified to fit each specific scenario, as explained below.

The architecture begins with a 7×7 convolutional layer followed by max-pooling. It includes four groups of residual blocks, each containing two blocks; these groups sequentially use 64, 128, 256, and 512 filters. Each residual block has two 3×3 convolutional layers, with identity shortcuts connecting the input directly to the output. The network concludes with a global pooling layer and a fully connected layer for classification, followed by a sigmoid activation function that outputs a single value between 0 and 1, allowing efficient learning and improved performance in deep learning tasks. The model is optimized using the stochastic gradient descent (SGD) optimizer, with categorical cross-entropy as the loss function, and is trained with a batch size of 128, ensuring stability and efficiency during training.

While the tested network architectures using only HST images or only LS images (see Fig. 7 images a and b, respectively) followed a straightforward approach based on the modified ResNet-18 architecture from previous studies, the innovation in this work lies in how we combine these data sets. For the stacked-images architecture (see Fig. 7c), we used cubic interpolation to convert the low-resolution LS images to high-resolution images matching those from the HST. This means that the final LS images had the same pixel resolution as the HST images after interpolation, without altering the intrinsic content of the original images. By doing this, we can directly use the four images (g, r, z, and F814W) together as input to the network architecture. In the training process, the input data was balanced to ensure that LS and HST contribute equally to the overall loss function. While the LS data set contains three bands (g, r, z) and HST only one band, each band in LS is treated as an independent input, effectively increasing the sample size of LS. To avoid bias, we applied a weighted loss function to balance the contribution of LS and HST data. Since each LS band (g, r, z) is treated as an independent image, making the LS dataset three times larger than the HST dataset, we assign a weight of 1 to HST samples and 1/3 to LS samples. This ensures equal contributions during training.

In the merged-branches scenario (see Fig. 7d), both types of astrophysical images (HST and LS) were processed through separate branches up to a point, but they ultimately contribute to a shared neural network. This includes a fully connected (FC) layer, meaning that while the initial layers handle the HST and LS data independently, the final layers merge to form a single network that processes both data types together. We then combined the outputs from these networks to carry out the final classification. To elaborate, each set of images is run through its dedicated network (HST images through a ResNet-18 model and LS images through another ResNet-18 model), with their feature maps flattened into vectors. We combine the flattened feature vectors to merge information from both networks into a new feature dimension. The input images of each object provided 512 features from the HST ResNet-18 and another 512 features from the LS model, resulting in a final feature vector of 1024 dimensions. This extensive feature set presented challenges like increased computational complexity, higher memory requirements, and potential overfitting. To tackle these challenges, we used another fully connected layer to reduce the 1024-dimensional feature vector to a lower dimension. A ReLU activation function was then applied to the output of the reduction layer. Next, the reduced feature vector undergoes a fully connected (dense) layer to refine the feature representation further. Finally, the processed feature vector was input into the output layer, where the sigmoid activation function is applied to yield a probability score ranging from 0 to 1 for each input image. A score close to 1 indicates a high likelihood of the image showing a gravitational lens, whereas a score near 0 suggests the absence of strong lensing, or other galaxy types rather than a lens.

Early stopping was applied to prevent overfitting, with a patience interval of 10 epochs; that is, the training stopped when the validation loss did not improve by more than 0.001 for 10 consecutive epochs. All four scenarios used the same training, validation, and test set (refer to Tab. 1). This allows us to draw a fair comparison of the performances for the four different scenarios.

4. Results and discussion

The loss curves for LS and HST alone are shown in Fig. 8 left panel. The training for LS required 304 epochs, while the training for HST was completed in 327 epochs. The evolution indicates that the networks (a) based on color images from LS train more quickly than those trained on single-band image of HST, which is expected due to the artifacts, different exposure times, and the higher resolution of HST that encodes substantial morphological information. A summary of the performance is shown in Table 2.

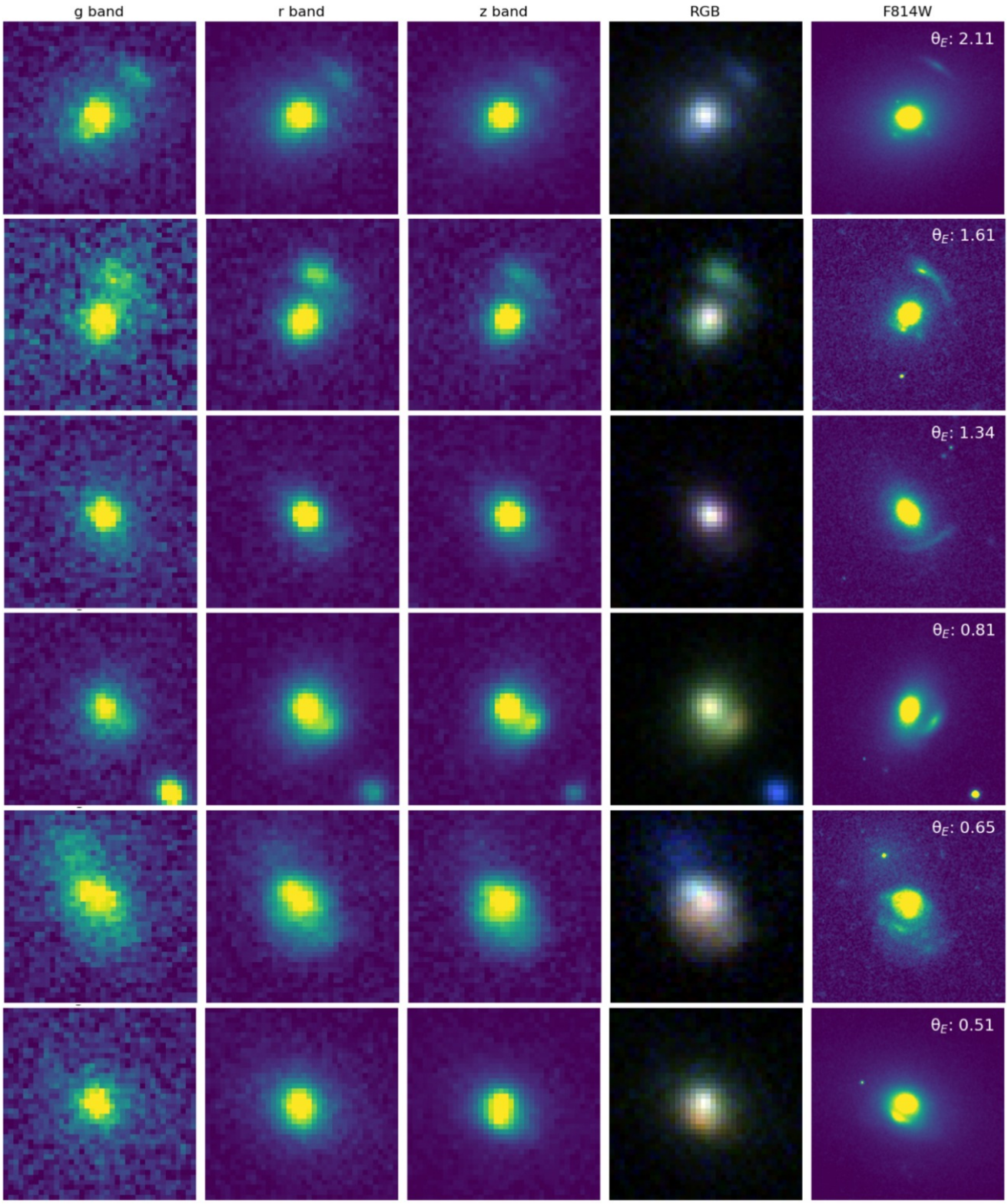

Fig. 6. Positive lens simulations of LS and HST for different Einstein radii (in descending order). From left to right: *g*, *r*, and *z* band images, plus the RGB image from LS. The last panel on the right shows the image in the F814W band from HST. Cutouts have a size of $10'' \times 10''$.

The stacked-images training loss (Fig. 8, right panel) shows a swift decrease, with the validation loss closely following the training loss but plateauing at a slightly higher value after around 100 epochs. This behavior suggests that the stacked-image architecture efficiently learns patterns from the training data, reaching stability relatively early but with a small generalization gap, as evidenced by the difference between the training and validation losses. The model has converged after 307 epochs, with both losses exhibiting minimal fluctuations after about 150 epochs. Contrarily, the merged-branch architecture exhibits a similar pattern, but the validation loss decreases more gradually than in the stacked-images model. The merged-branch loss continues

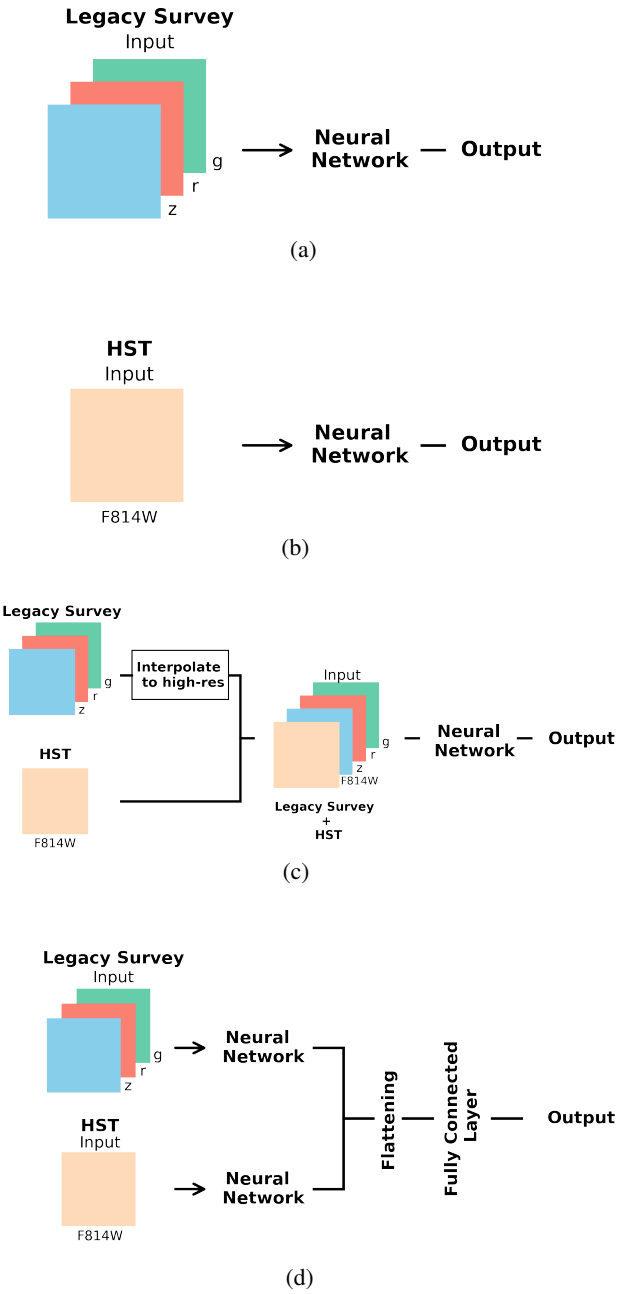

Fig. 7. Visualized architectures for the four scenarios: (a) LS, (b) HST architecture, (c) stacked-images architecture, and (d) merged-branches architecture.

to decrease past 150 epochs, converging after 321 epochs. This indicates that the merged architecture may require more training to fully integrate features from multiple branches, though it achieves a comparable final performance. The slower convergence could be due to the added complexity in combining independent networks. Despite this, the merged model still demonstrates competitive performance, particularly after fully integrating the features learned by the two branches.

Since we are more interested in the stacked-images architecture, we additionally created three more learning curve tests using different training and validation data (see Fig. 9). The first two training and their corresponding validation sets (Training loss 1, Training loss 2, Validation loss 1, and Validation loss 2) were calculated using randomly simulated data from HST and LS in two separate training and validation data sets. Subsequently, a third evaluation (Training loss 3 and Validation loss 3) incorporated the ∼360 real observational data of lens galaxy candidates identified in the literature to further refine the model's architecture. This test aimed to determine whether using only simulated data was sufficient for identifying new lenses. Importantly, the third validation set (Validation loss 3) demonstrated slightly better performance with a marginally lower loss than the first two validation sets derived from simulated data, though the improvement was not significant. Given this result, we opted to use only simulated data for training and reserved the real lens galaxy candidates for the test set.

After training, we measured the performance of the four scenarios in the receiver operating characteristic (ROC) curve, shown in Fig. 10 left panel. This curve illustrates the relation between the true-positive rate (TPR), which indicates the model's ability to identify actual positive cases correctly, and the false-positive rate (FPR), which measures the proportion of negative cases mistakenly classified as positives. The total number of negatives is approximately 120 000 in the test set. While these sample sizes allow us to quantify low FPR values, the estimates from HST and LS alone are less reliable in the low-FPR regime compared to the combined data set. This is primarily because HST provides high-resolution images that, while detailed, are noisier, and LS images, though offering broader wavelength coverage, suffer from lower resolution. At an FPR of 10^{-4}, the TPR is ∼0.41, ∼0.45, ∼0.51 and ∼0.55, for HST, LS, stacked images and merged-branches, respectively. The merged branches architecture achieves superior performance, maintaining a higher TPR at lower FPRs than the other architectures. This result underlines the potential benefits of integrating data from multiple telescopes to improve the accuracy of lens search detection. The performance of the HST and LS architectures alone can be compared, for instance, to the work done by Cañameras et al. (2021) with data from the Hyper Supreme Cam (HSC), which is comparable to what LSST data will be. Their work finds a TPR of ∼0.60, which is higher than our HST and LS TPRs (∼0.41 and ∼0.45, respectively). It is important to note, however, that their test set is constructed differently, as they used HSC Wide PDR2 images, while our study uses all available real galaxy-scale lens candidates found in the literature with corresponding HST and LS images. The image quality of HSC is expected to be comparable to that of LSST. In contrast, HST achieves a depth of 27.2 AB magnitude in the F814W filter for a single orbit (approximately 2028 seconds of exposure). The LS, conducted using the Dark Energy Camera Legacy Survey, reaches 5σ point source depths of approximately 24.7 AB magnitude in the g band, 23.9 AB magnitude in the r band, and 23 AB magnitude in the z band for areas with typical three-pass observations[5]. Accordingly, this comparison serves only as a reference point, with our primary conclusions drawn from the analysis of our four internal approaches rather than from the findings of Cañameras et al. (2021). We attribute the higher TPR in their results to the deeper and sharper image quality of HSC in comparison to LS, and the presence of multiple bands in comparison to HST.

In addition, to assess the significance of the performance difference between both architectures, we have made ten different ROC curves using ten different test sets (right panel in Fig. 10). Each test set is constructed by randomly splitting the overall data set to assess the robustness and performance of the architectures. We conduct this test to determine whether the merged-branches architecture performs better than the stacked architec-

[5] https://www.legacysurvey.org/decamls/

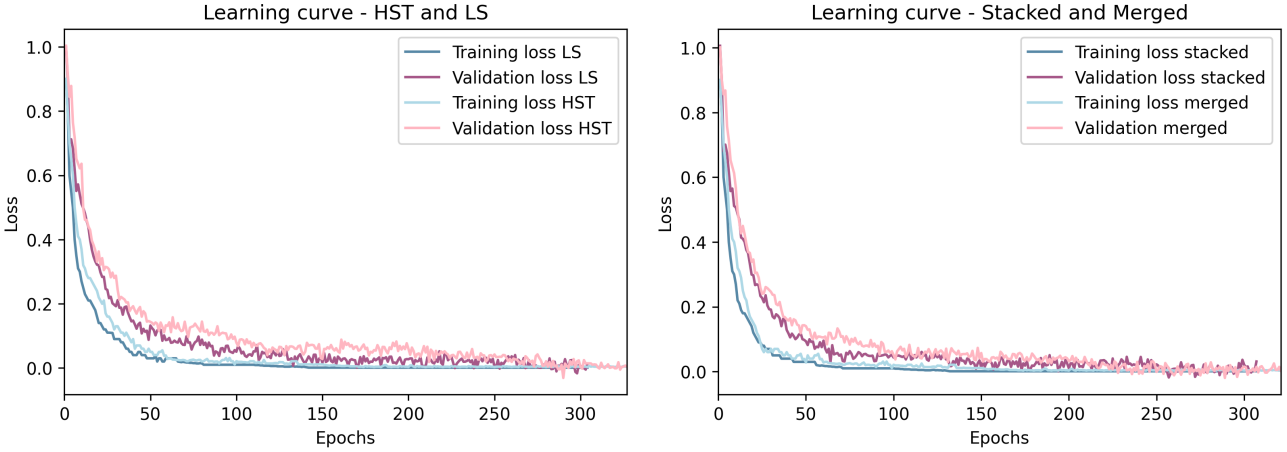

Fig. 8. Loss curves for the training (blue and light blue colors) and validation (purple and pink colors) for the four scenarios. Left: HST and LS alone. Right: stacked and merged branches.

Table 2. Summary of performance metrics for different training scenarios (see learning curves in Fig. 8).

Metric/Scenario	LS	HST	Stacked images	Merged branches
Epochs for Training	304	327	307	321
Learning Curve Characteristics	Fast initial learning with moderate stability post-training	Slower convergence due to noise, artifacts, and higher resolution	Swift decrease, with minimal fluctuations after 100 epochs; stable with small generalization gap	Gradual decrease in validation loss; stable with slower convergence due to added complexity
TPR at FPR of 10^{-4}	0.45	0.41	0.51	0.55

ture or is comparable in terms of its error. Based on the analysis of ten ROC curves, we can observe a higher performance of the merged-branches architecture. On the other hand, the lower performance in the stacked-images architecture could be attributed to the difficulty in processing the four images with two different resolutions/PSFs. This aspect will be further explored in a future paper to better interpret the network's decision-making process in classifying lens and non-lens images, examining how specific features influence its predictions and the factors contributing to potential misclassifications.

To further investigate the performance differences between the architectures, we performed an additional test by downsampling HST F814W images to match the resolution and pixel scale of the LS data set and incorporating them as a fourth LS filter. This modified LS network, using the four filters (g, r, z, and the downsampled F814W), was then compared to the four ROC curves mentioned previously (see Fig. 10). The results show that the performance of the "four-filter LS network" is comparable to that of the stacked-images architecture. This finding suggests that the stacked-images architecture does not fully exploit the high resolution of the HST images. The comparable performance between the two configurations indicates that the stacked-images network may struggle to integrate the additional information provided by higher-resolution data, potentially due to complexities introduced by varying resolutions. This limitation highlights the need for further investigation into how the architecture processes and combines multi-resolution data to improve its ability to extract meaningful features.

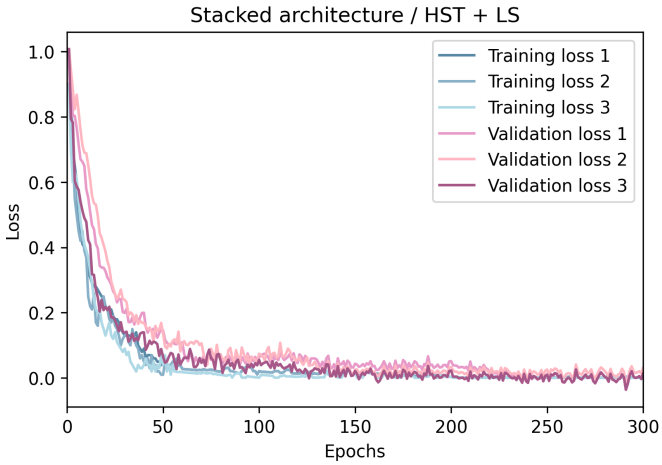

Fig. 9. Learning curves for the stacked-images architecture using three different HST and LS data sets. The data set that includes real lenses in the training (data set 3) has the best performance with the lowest loss.

Fig. 11 compares the CNN score distributions for the positive test set between the stacked-images and merged-branches approaches. The histograms display the normalized counts of CNN scores, segmented by the θ_E into two categories: $\theta_E \leq 1.3''$ (represented by filled histograms) and $\theta_E > 1.3''$ (represented by open histograms). We note that 65% of the positive mock have $\theta_E < 1.3''$. For the category $\theta_E > 1.3''$, both models show a significant increase in CNN scores close to 1.0, reflecting the two

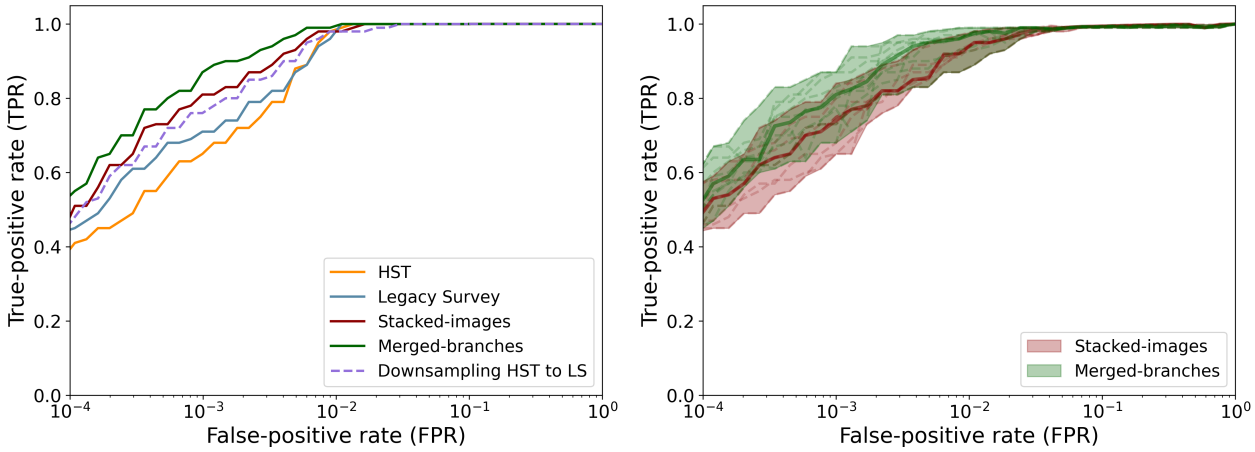

Fig. 10. ROC curves of different network architectures. Left panel: our four main networks illustrated in Fig. 7 (solid lines), and the network that downsampled HST F814W to LS (dashed purple line). Right panel: ten different test sets for the stacked-images and merged-branches network, illustrating that the merged-branches network overall performs better than the stacked-images network.

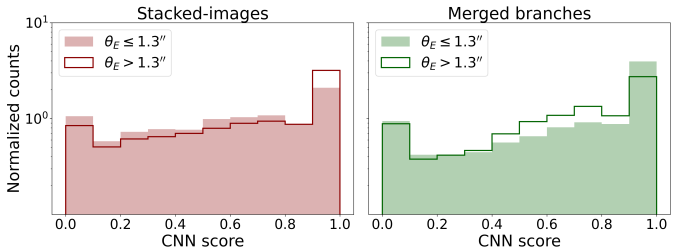

Fig. 11. CNN scores for the positive examples of the test set.

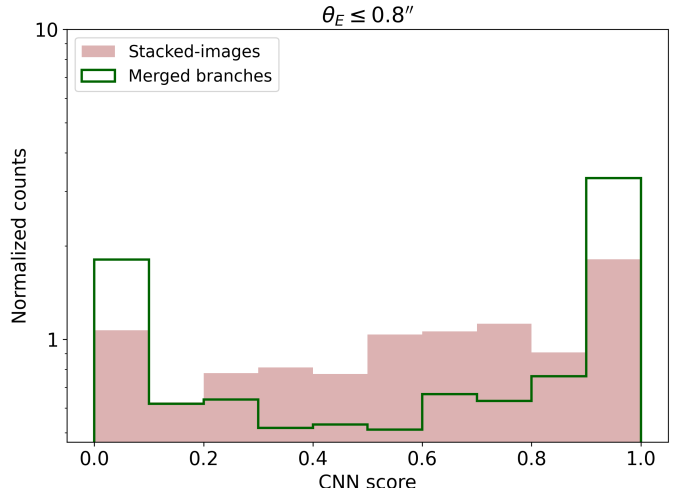

Fig. 12. CNN scores for systems with $\theta_E \leq 0.8''$ from the test set scored by both architectures.

architectures' ability to confidently identify these examples as positive.

In the merged-branches model (right panel), the CNN scores for $\theta_E \leq 1.3''$ shows a significant peak at a CNN score of 1.0. The final bin (with scores between 0.9 and 1.0) contains 2445 mock lenses (25% of the total positive test set), indicating that the model frequently assigns the highest confidence to these cases. Notably, 1856 examples in this category achieved a CNN score of precisely 1.0, highlighting the model's strong tendency toward high-certainty classifications. However, for $\theta_E > 1.3''$, it displays a less pronounced increase in CNN scores near 1.0. This suggests that, despite a subset of confident classifications, the model is generally less reliable in distinguishing lenses with larger θ_E values.

On the other hand, the stacked-images branch (left panel of Fig. 11) shows a lower peak at 1.0 for $\theta_E \leq 1.3''$, with 1208 mock lenses in the final bin. The model shows 323 examples with a CNN score of precisely 1.0. For $\theta_E > 1.3''$, the stacked-images model displays 885 examples reaching a score of 1.0. However, the generally lower peaks near scores of 0.9–1 suggest that the stacked-images model may perform less confidently for lower θ_E values compared to the merged-branches model.

Figure 12 compares the performance of the stacked-images and merged-branches models on systems with low Einstein radii ($\theta_E \leq 0.8''$). The stacked-images model shows a gradual increase in CNN scores, with counts spread across bins and peaking at 1.0. This indicates a progressive assignment of confidence, with varying confidence levels before reaching the highest score. In contrast, the merged-branches model exhibits sharper increases at specific score bins, with a strong concentration of scores at 1.0. This suggests that the merged-branches model tends to assign high-confidence scores more directly for systems it identifies as lenses. While both models peak at 1.0, the merged-branches model does so more abruptly, highlighting its preference for high-confidence classifications for systems with $\theta_E \leq 0.8''$.

Given the total of approximately 8000 positive examples (of which 364 are real lens candidates), the merged branches model's tendency to assign extreme confidence scores may be beneficial in scenarios prioritizing high-confidence predictions for low Einstein radius lenses. However, this focus comes at the potential cost of flexibility across a broader score distribution, as seen in the stacked-images model, which maintains a more even distribution across scores, accommodating systems with both high and low confidence levels.

Fig. 13 shows a selection of very likely strong lenses systems, along with the model networks' confidence in detecting lensing features: (a) LS-only, (b) HST-only, (c) stacked-images, and (d) merged-branches. The HST-only model generally provides higher scores compared to the LS-only model on positive examples. This is due to the higher resolution and clearer features in the HST images, making lensing features such as arcs and multiple images more distinguishable. However, the best performance across all cases is seen in the merged model (case d), followed closely by the stacked model (case c).

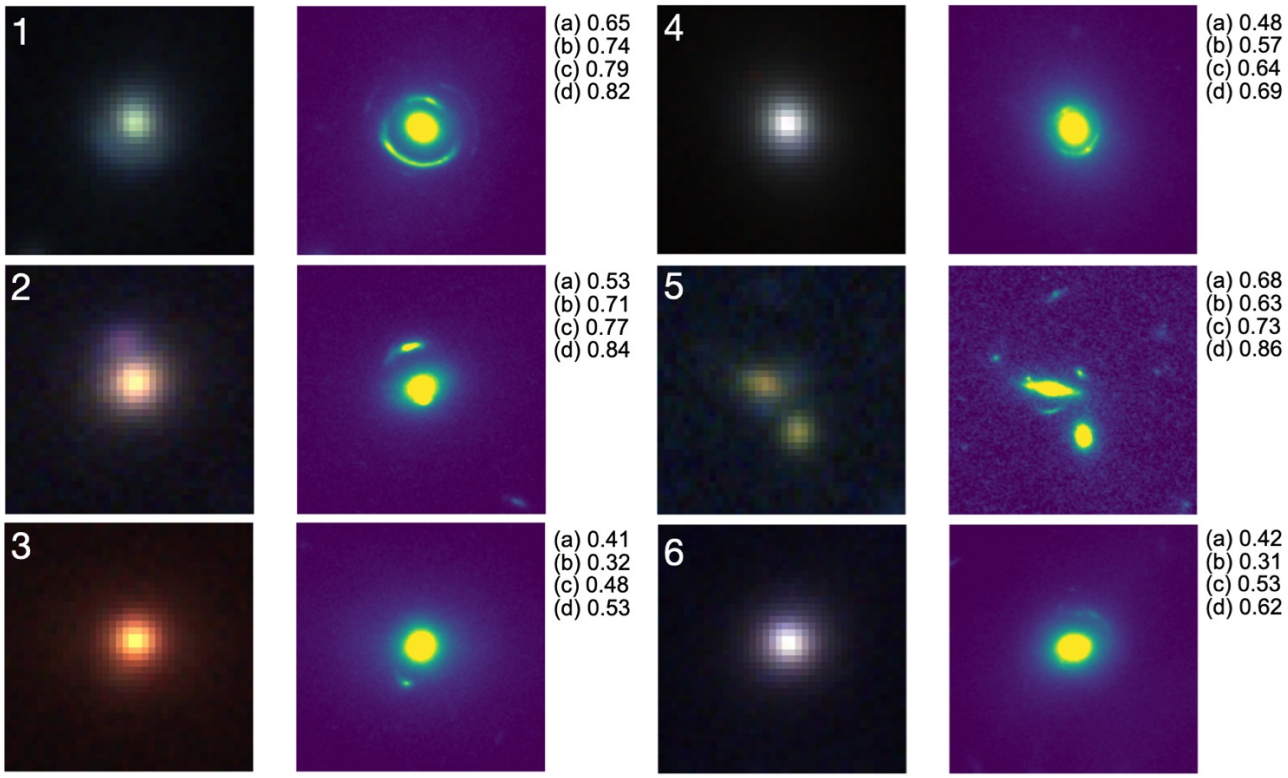

Fig. 13. CNN scores of some test set examples of gravitational lensing (Bolton et al. 2008; Pawase et al. 2014) for LS (RGB image) and HST (F814W). The letters a), b), c), and d) represent the scores from the LS-only, HST-only, stacked-images, and merged-branches network architectures.

In cases where clear lensing features, such as arcs and multiple images, are visible (e.g., examples 1, 2, and 5), the scores are higher across all models, particularly in the merged and stacked models. For instance, in the 5th image, the scores for the merged model (d) and stacked model (c) reach 0.86 and 0.73, respectively, indicating a high level of confidence in detecting a lensing feature. The LS-only model (a) also performs well in this case, with a score of 0.68, but it does not reach the level of accuracy achieved by the combined models.

Conversely, for images without obvious lensing features (e.g., examples 3 and 6), even the merged and stacked models show lower confidence. This demonstrates that while combining data from different surveys improves detection, strong and clear lensing signals are still crucial for high-confidence predictions. The LS-only model (a) performs particularly poorly in these cases, with scores of 0.41 in case 3 and 0.42 in case 6, further highlighting the limitations of low-resolution data when clear features are absent.

5. Conclusions

In this paper, we have introduced a novel architecture that combines high-resolution data from HST with multiband, lower-resolution data from LS to improve gravitational lens detection. The first model uses only single-band, high-resolution HST images, capturing fine structural details but lacking the multiband information available in other configurations. The second model relies solely on multiband LS images, which provide broad coverage across three bands but at a lower resolution. The stacked architecture combines four images (g, r, z filters of LS and F814W filter of HST) by interpolating LS images to match the pixel scale of HST and processing them simultaneously. Finally, the merged-branches architecture processes HST and LS data separately in independent branches before merging their outputs for classification.

Overall, the merged-branches architecture for HST and LS data significantly boosts detection rates in many cases, underscoring the strength of multi-source data fusion. Although multiband high-resolution data would be ideal, such data sets are not yet available for large sky areas. However, our approach shows that combining high-resolution single-band data with multiband, low-resolution images greatly enhances detection capabilities. This indicates that future lensing searches could benefit from similar multi-source data integration until multiband high-resolution data become more widely available.

This methodology offers great potential for integrating diverse data sets, and the proposed pipeline can be applied to various types of imaging, beyond Euclid and LSST, and is not restricted to data from differing resolutions. Looking ahead, applying these networks to Euclid and LSST images is feasible within the next few months or years, though further development and validation are necessary to confirm this potential. For Euclid, we expect even better performance from the combination of images from different instruments, such as the high-resolution VIS images and the NISP bands, allowing for a richer data set. This approach, similar to our work of combining images of different resolution, should further improve the detection capabilities compared to HST and LS alone. Additionally, LSST's multiband data will provide complementary coverage, further enhancing performance.

Acknowledgements. SHS thanks the Max Planck Society for support through the Max Planck Fellowship. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research

and innovation programme (grant agreement No 771776). This research is supported in part by the Excellence Cluster ORIGINS which is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC-2094 – 390783311. SS has received funding from the European Union's Horizon 2022 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 101105167 – FASTIDIoUS. SB acknowledges the funding provided by the Alexander von Humboldt Foundation.

References

Acevedo Barroso, J. A., O'Riordan, C. M., Clément, B., et al. 2025, A&A, 697, A14
Arendse, N., Dhawan, S., Sagués Carracedo, A., et al. 2024, MNRAS, 531, 3509
Auger, M. W., Treu, T., Bolton, A. S., et al. 2009, ApJ, 705, 1099
Bag, S., Huber, S., Suyu, S. H., et al. 2024, A&A, 691, A100
Beckwith, S. V. W., Stiavelli, M., Koekemoer, A. M., et al. 2006, AJ, 132, 1729
Bolton, A. S., Burles, S., Koopmans, L. V. E., et al. 2008, ApJ, 682, 964
Bolton, A. S., Burles, S., Koopmans, L. V. E., Treu, T., & Moustakas, L. A. 2006, ApJ, 638, 703
Bom, C. R., Fraga, B. M. O., Dias, L. O., et al. 2022, MNRAS, 515, 5121
Brownstein, J. R., Bolton, A. S., Schlegel, D. J., et al. 2012, ApJ, 744, 41
Cañameras, R., Schuldt, S., Suyu, S. H., et al. 2020, A&A, 644, A163
Cañameras, R., Schuldt, S., Shu, Y., et al. 2021, A&A, 653, L6
Cañameras, R., Schuldt, S., Shu, Y., et al. 2024, A&A, 692, A72
Capak, P., Aussel, H., Ajiki, M., et al. 2007, ApJS, 172, 99
Capak, P., Cuillandre, J. C., Bernardeau, F., et al. 2019, ArXiv e-prints [arXiv:1904.10439]
Chan, J. H. H., Suyu, S. H., Sonnenfeld, A., et al. 2020, A&A, 636, A87
Chan, J. H. H., Wong, K. C., Ding, X., et al. 2024, MNRAS, 527, 6253
Chen, W., Kelly, P. L., Oguri, M., et al. 2022, Nature, 611, 256
Collett, T. E. 2015, ApJ, 811, 20
Diehl, H. T., Buckley-Geer, E. J., Lindgren, K. A., et al. 2017, ApJS, 232, 15
Euclid Collaboration (Mellier, Y., et al.) 2025, A&A, 697, A1
Faure, C., Kneib, J.-P., Covone, G., et al. 2008, ApJS, 176, 19
Frye, B. L., Pascale, M., Pierel, J., et al. 2024, ApJ, 961, 171
Garvin, E. O., Kruk, S., Cornen, C., et al. 2022, A&A, 667, A141
Goldstein, D. A., Nugent, P. E., & Goobar, A. 2019, ApJS, 243, 6
Goobar, A., Amanullah, R., Kulkarni, S. R., et al. 2017, Science, 356, 291
Goobar, A. A., Johansson, J., Dhawan, S., et al. 2022, Transient Name Server AstroNote, 180, 1
He, K., Zhang, X., Ren, S., & Sun, J. 2016, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1
Huang, X., Storfer, C., Ravi, V., et al. 2020, ApJ, 894, 78
Inami, H., Bacon, R., Brinchmann, J., et al. 2017, A&A, 608, A2
Ivezic, Z., Axelrod, T., Brandt, W. N., et al. 2008, Serb. Astron. J., 176, 1
Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, ApJ, 873, 111
Jacobs, C., Collett, T., Glazebrook, K., et al. 2019, ApJS, 243, 17
Jaelani, A. T., More, A., Oguri, M., et al. 2020, MNRAS, 495, 1291
Jaelani, A. T., More, A., Wong, K. C., et al. 2024, MNRAS, 535, 1625
Kassiola, A., & Kovner, I. 1993, ApJ, 417, 450
Kelly, P., Zitrin, A., Oguri, M., et al. 2022, Transient Name Server AstroNote, 169, 1
Kelly, P. L., Rodney, S. A., Treu, T., et al. 2015, Science, 347, 1123
Krist, J. 1995, in Astronomical Data Analysis Software and Systems IV, eds. R. A. Shaw, H. E. Payne, & J. J. E. Hayes, ASP Conf. Ser., 77, 349
Lanusse, F., Ma, Q., Li, N., et al. 2018, MNRAS, 473, 3895
Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, ArXiv e-prints [arXiv:1110.3193]
Metcalf, R. B., Meneghetti, M., Avestruz, C., et al. 2019, A&A, 625, A119
Myers, S. T., Jackson, N. J., Browne, I. W. A., et al. 2003, MNRAS, 341, 1
Nikutta, R., Fitzpatrick, M., Scott, A., & Weaver, B. A. 2020, Astron. Comput., 33, 100411
Oguri, M., & Marshall, P. J. 2010, MNRAS, 405, 2579
Oguri, M., Inada, N., Pindor, B., et al. 2006, AJ, 132, 999
Pawase, R. S., Courbin, F., Faure, C., Kokotanekova, R., & Meylan, G. 2014, MNRAS, 439, 3392
Petrillo, C. E., Tortora, C., Vernardos, G., et al. 2019, MNRAS, 484, 3879
Pierel, J. D. R., Newman, A. B., Dhawan, S., et al. 2024, ApJ, 967, L37
Polletta, M., Nonino, M., Frye, B., et al. 2023, A&A, 675, L4
Pourrahmani, M., Nayyeri, H., & Cooray, A. 2018, ApJ, 856, 68
Refsdal, S. 1964, MNRAS, 128, 307
Rodney, S. A., Brammer, G. B., Pierel, J. D. R., et al. 2021, Nat. Astron., 5, 1118
Rojas, K., Savary, E., Clément, B., et al. 2022, A&A, 668, A73
Sainz de Murieta, A., Collett, T. E., Magee, M. R., et al. 2024, MNRAS, 535, 2523
Schaefer, C., Geiger, M., Kuntzer, T., & Kneib, J. P. 2018, A&A, 611, A2
Schneider, P., Ehlers, J., & Falco, E. E. 1992, Gravitational Lenses
Schuldt, S., Suyu, S. H., Meinhardt, T., et al. 2021, A&A, 646, A126
Schuldt, S., Cañameras, R., Andika, I. T., et al. 2025, A&A, 693, A291
Scoville, N., Abraham, R. G., Aussel, H., et al. 2007, ApJS, 172, 38
Shajib, A. J., Vernardos, G., Collett, T. E., et al. 2024, Space Sci. Rev., 220, 87
Shu, Y., Cañameras, R., Schuldt, S., et al. 2022, A&A, 662, A4
Sonnenfeld, A., Chan, J. H. H., Shu, Y., et al. 2018, PASJ, 70, S29
Storfer, C., Huang, X., Gu, A., et al. 2024, ApJS, 274, 16
Suyu, S. H., & Halkola, A. 2010, A&A, 524, A94
Suyu, S. H., Hensel, S. W., McKean, J. P., et al. 2012, ApJ, 750, 10
Suyu, S. H., Huber, S., Cañameras, R., et al. 2020, A&A, 644, A162
Treu, T. 2010, ARA&A, 48, 87
Turner, E. L., Ostriker, J. P., & Gott, J. R., I. 1984, ApJ, 284, 1
Walmsley, M., Lintott, C., Géron, T., et al. 2022, MNRAS, 509, 3966
Wilde, J., Serjeant, S., Bromley, J. M., et al. 2022, MNRAS, 512, 3464
Willett, K. W., Galloway, M. A., Bamford, S. P., et al. 2017, MNRAS, 464, 4176
Wojtak, R., Hjorth, J., & Gall, C. 2019, MNRAS, 487, 3342
Wong, K. C., Chan, J. H. H., Chao, D. C. Y., et al. 2022, PASJ, 74, 1209
Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, AJ, 140, 1868
Zaborowski, E. A., Drlica-Wagner, A., Ashmead, F., et al. 2023, ApJ, 954, 68
Zhou, R., Dey, B., Newman, J. A., et al. 2023, AJ, 165, 58