ViT-based Local Volume Dwarf Galaxy Identification (VIDA) in the CSST survey

Han Qu, Zhen Yuan, Chengliang Wei, Chao Liu, Jiang Chang, Guoliang Li, Nicolas F Martin, Chaowei Tsai, Shi Shao, Yu Luo, et al.

Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY
MNRAS **544,** 1238–1254 (2025) https://doi.org/10.1093/mnras/staf1586

ViT-based Local Volume Dwarf Galaxy Identification (VIDA) in the CSST survey

Han Qu(曲涵),[1,2] Zhen Yuan(袁珍),[3★] Chengliang Wei(韦成亮),[1] Chao Liu(刘超),[4,5] Jiang Chang(常江),[1] Guoliang Li(李国亮),[1] Nicolas F. Martin,[6,7] Chaowei Tsai(蔡肇伟),[4,8,9] Shi Shao(实),[4] Yu Luo(罗煜),[10] Ran Li(李然),[11] Xi Kang(康熙),[12] Xiangxiang Xue(薛香香)[4] and Zhou Fan(范舟)[4]

[1]*Purple Mountain Observatory, Chinese Academy of Sciences, Nanjing 210008, China*
[2]*School of Astronomy and Space Sciences, University of Science and Technology of China, Hefei 230026, China*
[3]*School of Astronomy and Space Science, Nanjing University, Nanjing 210093, China*
[4]*National Astronomical Observatories, Chinese Academy of Sciences, 20A Datun Road, Chaoyang District, Beijing 100012, China*
[5]*Research Center for Astronomical Computing, Zhejiang Laboratory, Hangzhou 311121, China*
[6]*Observatoire Astronomique de Strasbourg, Université de Strasbourg, CNRS, UMR 7550, F-67000 Strasbourg, France*
[7]*Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany*
[8]*Institute for Frontiers in Astronomy and Astrophysics, Beijing Normal University, Beijing 102206, China*
[9]*School of Astronomy and Space Science, University of Chinese Academy of Sciences, Beijing 100049, China*
[10]*School of Physics and Information Science, Hunan Normal University, Changsha 410081, China*
[11]*Department of Astronomy, Beijing Normal University, Beijing 100875, China*
[12]*Zhejiang University-Purple Mountain Observatory Joint Research Center for Astronomy, Zhejiang University, Hangzhou 310027, China*

Accepted 2025 September 12. Received 2025 September 10; in original form 2025 June 3

ABSTRACT

Identifying dwarf galaxies within the Local Volume is crucial for constraining the luminosity function of satellite galaxies in the nearby universe. We report the detection capabilities of dwarf galaxies within the Local Volume using the Chinese Space Station Telescope (CSST). Based on the simulated imaging data of CSST, we present VIDA, a ViT-based dwarf galaxy identification Algorithm designed for detecting Local Volume dwarf galaxies. The simulated Local Volume dwarf galaxies can be identified using a pre-processing method for 'extended source detection', followed by classification with a pretrained ViT-Base model. This pipeline achieves a true positive rate exceeding 85 per cent with a false positive rate of only 0.1 per cent. We quantify the detection completeness of Local Volume dwarf galaxies across a three-dimensional parameter space defined by absolute magnitude (M_V), half-light radius (R_h), and heliocentric distance, based on simulated single-exposure CSST wide-field imaging survey data. For unresolved or semiresolved dwarf galaxies, our method achieves a significantly deeper absolute magnitude detection limit compared to catalogue-based approaches, reaching $M_V = -7$ within 10 Mpc with a surface brightness threshold $\mu \sim 25$ mag/arcsec^2 at 2–5 Mpc and ∼26 mag/arcsec^2 at 5–10 Mpc. While traditional matched-filter techniques based on stellar catalogues remain more effective for detecting fully resolved, extremely low surface brightness galaxies within 5 Mpc, our approach offers complementary strengths – particularly in identifying compact or more distant systems – making it a valuable tool for expanding the census of Local Volume dwarf galaxies.

Key words: galaxies: dwarf.

1 INTRODUCTION

The faint end of satellite galaxy luminosity functions are sensitive to cosmological models with different dark matter properties (Governato et al. 2015; Dekker et al. 2022; Forouhar Moreno et al. 2022). The satellite galaxies around the Milky Way (MW) and the Andromeda galaxy (M31) are often studied for their satellite luminosity functions (Koposov et al. 2008; Martin et al. 2016; Doliva-Dolinsky et al. 2023; Homma et al. 2024; Doliva-Dolinsky, Collins & Martin 2026,). Constraining cosmological models on small scales requires a well-surveyed dwarf galaxy population below the current detection limit ($M_V > -5$). It is also important to broaden satellite luminosity function investigations beyond the Local Group to the Local Volume (∼20 Mpc) to encompass a statistically significant sample of galaxy systems. Substantial progress has been made with recent surveys (Bennet et al. 2020; Carlsten et al. 2020; Engler et al. 2021; Davis et al. 2021a; Gozman et al. 2024; Kanehisa et al. 2024), such as the Local Volume Legacy survey (Lee et al. 2008), the Dragonfly Nearby Galaxies survey (Merritt et al. 2016), the Exploration of Local VolumE Satellites (ELVES)

★ E-mail: zhen.yuan@nju.edu.cn

survey (Carlsten et al. 2022), and the SAGA survey (Tollerud et al. 2022). So far, the observed faint end of the satellite luminosity in the Local Volume has been pushed to $M_V < -9$ (Crosby et al. 2023).

The fourth-generation of large-scale survey telescopes, such as the Legacy Survey of Space and Time (Ivezić et al. 2019), the *Euclid Space Telescope* (*Euclid*; Euclid Collaboration 2022; Cuillandre et al. 2025), and the Chinese Space Station Telescope (CSST; Zhan 2021) will provide deeper and wider-area imaging survey data. These advance can significantly improve satellite galaxy searches within the Local Volume. The 2-m CSST, set to launch in the near future, features a main survey camera with a wide 1.1 deg^2 field-of-view and a spatial resolution of ∼0.15 arcmin. It covers near-ultraviolet to near-infrared wavelengths with NUV and u, g, r, i, z, y filters. The limiting magnitude in the r band reaches 26 mag in CSST's main survey. All these capabilities would make CSST a powerful tool to detect dwarf galaxies in the Local Volume.

Two methods for searching for nearby dwarf galaxies are often used: catalogue-based (Koposov et al. 2008; Walsh, Willman & Jerjen 2009) and image-based methods (Carlsten et al. 2020; Davis et al. 2021b). For dwarf galaxies within 10 Mpc , their member stars are spatially resolvable with space telescopes, making catalogue-based methods suitable. The overall procedure of this approach is to first select stars along the giant branch on colour–magnitude diagram. Then, the significance of the overdensity of these selected stars are evaluated, which allows for the identification of dwarf galaxy candidates. Classic methods such as the 'matched filter' technique (Drlica-Wagner et al. 2015; Laevens et al. 2015; Simon 2019) and those based on likelihood estimators (Martin et al. 2013) have led to successful discoveries of Local Group dwarf galaxies. In our previous work, we utilize the classic approach and evaluated the detection limits of dwarf galaxies with the CSST (Qu et al. 2023, Qu23 hereafter), achieving a limiting absolute magnitude of $M_V = -5.8$ and a surface brightness threshold of $\mu_{250} = 29.7\,\mathrm{mag\,arcsec^{-2}}$ within the 1–2 Mpc range. For more distant dwarf galaxies in the Local Volume up to 20 Mpc , stars at their centres are hard to resolve with current instrumentations. Besides, there are fewer stars brighter than the limiting magnitude of CSST wide survey, making the catalogue-based searches less effective beyond 5 Mpc . In addition, stars at the centres of dwarf galaxies may not be individually resolved. The integrated light from a few bright giant stars, combined with the diffuse contribution of numerous faint member stars, causes these galaxies to appear more like extended sources. As a result, some of these dwarf galaxies remain visually detectable in imaging data (Bennet et al. 2017; Danieli, van Dokkum & Conroy 2018). To fully leverage the information contained in these images, we present a novel image-based search method designed to enhance the detectability of distant dwarf galaxies within the Local Volume ('LV dwarf galaxies' afterward).

A typical approach adopted in existing studies is to first detect extended sources from images by setting thresholds for signal to noise ratio (S/N) and angular size. Candidate sources are then validated through visual inspection (Bennet et al. 2017; Carlsten et al. 2020). While practical for searching satellite galaxies around individual host galaxies, this approach becomes prohibitively inefficient for large-area surveys with the CSST and future imaging survey experiments. These large data sets demand fully automated pipelines to screen candidates, as manual inspections are unsustainable. Fortunately, the distinguishing features of LV dwarf galaxies, traditionally leveraged by human classifiers to separate them from distant galaxies or galaxy groups, are inherently compatible with modern image recognition models.

Image recognition models have been widely applied to identify specific astronomical objects, such as classifications of gravitational lensing systems (Shu et al. 2022), galaxy morphologies (Robertson et al. 2023; Fernández-Iglesias, Buitrago & Sahelices 2024) and the detection of nearby dwarf galaxies and low surface brightness systems (Zaritsky et al. 2019; Tanoglidis et al. 2021; Jones et al. 2024). Recent advancements in machine learning have led to the development of highly transferable models (Bhavanam et al. 2024), which achieve high performance with minimal computational effort through fine-tuning of pre-trained architectures. Among these, the Vision Transformer (ViT) – a transformer-based model originally designed for natural image processing – has emerged as a powerful tool. In astronomy, ViT has demonstrated superiority over traditional Convolutional Neural Networks (CNNs) across a range of tasks, including galaxy morphology classification (Lin et al. 2021), gravitational lensing detection (Huang et al. 2022), and cosmological parameter inference (Gondhalekar & Moriwaki 2024).

This work constitutes the second paper in our series on nearby dwarf galaxy detection with the CSST, extending the scope of our previous study (Qu23), which focused on systems within 1 Mpc. Here, we expand the search radius to 20 Mpc. The core of our approach is utilizing a transformer-based image recognition ViT model to improve the CSST's ability to detect LV dwarf galaxies over a broader distance range and to better identify faint, small systems.

This paper is organized as follows: in Section 2, we introduce the CSST image simulation pipeline, including the fiducial input catalogues that contain artificial LV dwarf galaxies and the CSST Image Simulator. The image-based detection method, considering different background galaxies in fiducial catalogues and its performance is described in Section 3 and Section 4.1. We discuss the post-processing in Section 5. The summary is provided in Section 6.

2 MOCK IMAGES

The construction of realistic mock images as close to the real observations as possible is the first and a crucial step in our work. This process starts with a fiducial catalogue that contains galaxies and stars. Then a set of artifical LV dwarf galaxies are injected into it. This mock catalogue is input into the CSST Image Simulator (Wei et al., in preparation) to produce the mock CSST images.

2.1 Fiducial catalogue

The fiducial catalogue consists of two components: a synthetic MW stellar population generated using the population synthesis model *TRILEGAL* (Girardi 2016), which produces mock stellar catalogues based on the Galactic structure components (thin disc, thick disc, halo, and bulge) while accounting for the extinction, photometric systems, and star formation histories; and galaxies from the cosmological simulation suite *Jiutian-1G* (Han et al. 2025,), referred to as background galaxies in this work (see Wei et al., in preparation). *Jiutian-1G* is a high-resolution dark matter-only N-body cosmological simulation performed using the LGadget-3 code, adopting the Planck 2018 cosmological parameters (Planck Collaboration VI 2020). The simulation spans a comoving box of $1000\,h^{-1}$ Mpc per side with 6144^3 particles and achieves a mass resolution of $3.723 \times 10^8\,\mathrm{M}_\odot$. Galaxies are populated using both semi-analytical models (e.g. GAEA, LGalaxies) and subhalo abundance matching. Based on this simulation, a mock light-cone galaxy catalogue extending to redshift $z \sim 3.5$ is constructed. A full-sky ray-tracing simulation is conducted to obtain weak gravitational lensing signals, including shear and magnification, at each galaxy's

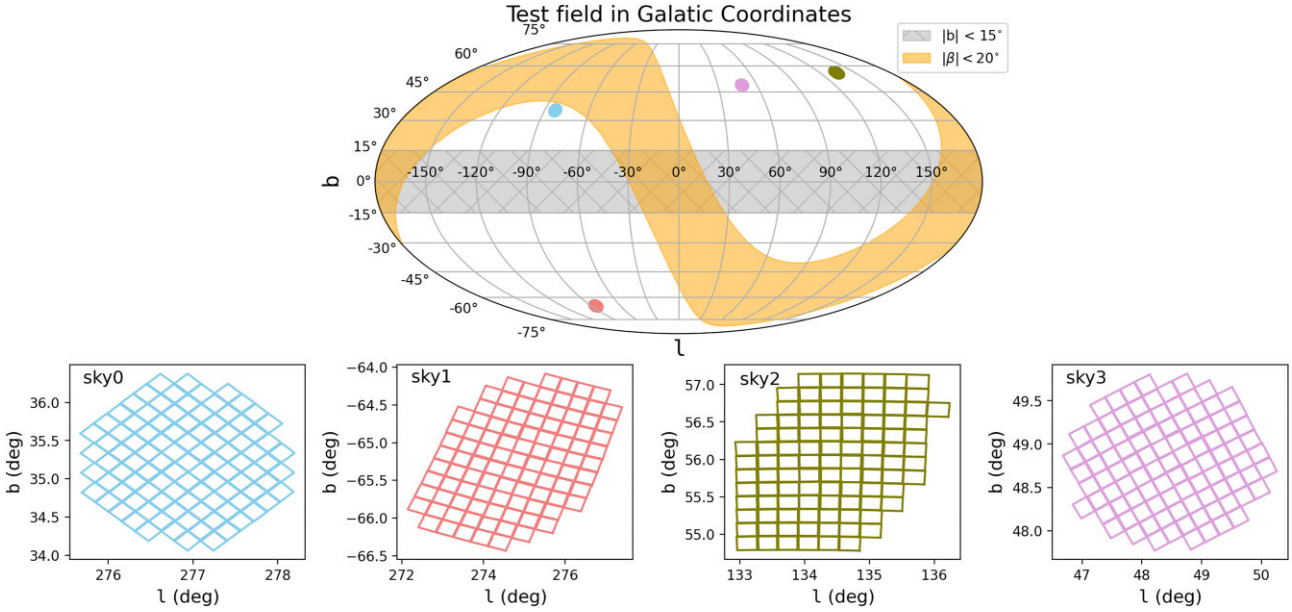

Figure 1. Distribution of the simulated sky regions in Galactic coordinates. The grey (hatched) and orange regions correspond to$|b| < 15^\circ$ and$|\beta| < 20^\circ$, respectively, which are areas not covered by the main CSST survey.

position. The final galaxy catalogue includes positions, redshifts, stellar masses, morphologies, sizes, spectral energy distributions (SEDs), shear, and magnification.

As shown in Fig. 1, our fiducial catalogue comes from four sky regions, which represent typical CSST regions by avoiding $|b| < 15^\circ$ (galactic latitude) and $|\beta| < 20^\circ$ (ecliptic latitude). Every sky region includes 100 mosaicking patches that corresponds to the projected CCD footprint of a single chip covering $11.'4 \times 11.'4$ on the sky. We generate source catalogues in three photometric bands: g, r, and i. The distance distribution of the MW stars as well as their luminosity function in the r-band magnitudes of each sky region is shown as solid histograms in the upper row of Fig. 2. Similarly, the redshift distribution and the luminosity function of the background galaxies are shown in the bottom row. The faint end of the luminosity function of both stars and galaxies are well below the depth ($r \approx 25.5$ mag) of the CSST survey for a single exposure of 150s.

2.2 Artificial Local Volume dwarf galaxies

Following the recipe presented in Qu23, artificial LV dwarf galaxies are constructed using single stellar population PARSEC models[1] (Bressan et al. 2012). The stellar populations of nearby dwarf galaxies are predominantly old and metal-poor. While some nearby dwarfs exhibit recent star formation, in our simulations we adopt a simplified set-up (uniform age of 11 Gyr and metallicity of [M/H] $= -2.0$), reflecting this general characteristic. This choice is sufficient for testing the performance of our detection algorithm under standard conditions. The stellar radial density profiles are modelled with an exponential profile. The LV dwarf galaxies have stellar mass distributions from 10^3 to 10^6 $M_\odot$ and half-light radii from 10 to 316 pc , see Table 1. The distance range is set from 316 kpc to 20 Mpc , which extends the exploration from the Local Group to the Local Volume. Given that currently known LV dwarf galaxies within the Local Group are almost complete with stellar masses above 10^5 $M_\odot$ (Drlica-Wagner et al. 2021; Doliva-Dolinsky et al. 2023), we set a lower distance limit of 1 Mpc when testing these relatively massive systems. In total, 1953 artificial LV dwarf galaxies are generated. Most of the simulated dwarfs we tested have sizes smaller than the typical LV dwarf galaxies (Danieli et al. 2018). However, they correspond to the relatively smaller-size dwarfs observed in current surveys. Our subsequent results indicate that the algorithm in this work shows particular advantages in detecting this class of compact LV dwarf galaxies.

The stellar catalogue for each LV dwarf galaxies is constructed using the catalogue of its member stars, enabling a realistic simulation of the light emission processes of nearby dwarf galaxies. For each simulated LV dwarf galaxy, the original star catalogue included all theoretical members provided by the PARSEC models. We then truncated the catalogues based on distance: dwarfs within 1 Mpc retain the brightest 8 per cent of member stars, those between 1 and 3 Mpc retain the brightest 6 per cent, and dwarfs beyond 3 Mpc retain the brightest 4 per cent. These cuts are well below the CSST point-source detection limit for instance, the stellar magnitude cut is ∼29 mag in g-band for a 316 kpc dwarf galaxy and ∼40 mag for a 20 Mpc dwarf galaxy. This approach removes extremely faint stars that contribute negligibly to the simulated images while keeping enough relatively faint members; tests show that the total photon loss compared to the full catalogue is less than 0.1 per cent, ensuring that the simulated images closely approximate reality.

2.3 CSST Image Simulator

Mock images are generated using the CSST Image Simulator[2], which is developed by the CSST scientific data processing and analysis system. By combining the mock galaxy catalogue from cosmological simulations with weak gravitational lensing effects and detailed instrument modelling, the simulator is aimed to produce realistic

[1] http://stev.oapd.inaf.it/cgi-bin/cmd

[2] https://csst-tb.bao.ac.cn/code/csst_sim/csst-simulation

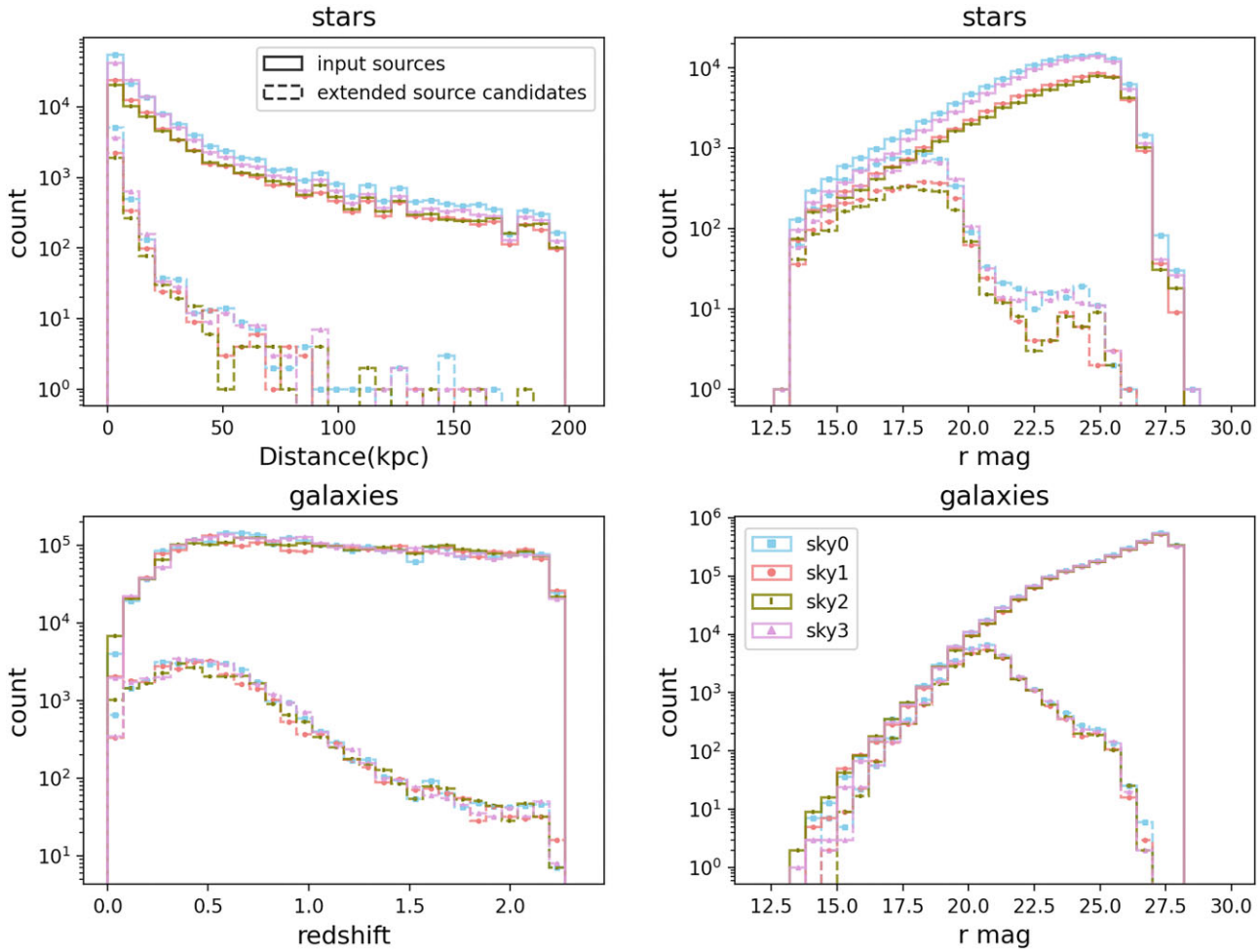

Figure 2. The distance (redshift) and apparent magnitude (in the *r* band) distributions of stars and background galaxies in the fiducial catalogue for the simulation programme across our four test sky regions. Different colours with distinct markers represent data from different sky regions, with solid lines indicating the input catalogue and dashed lines representing sources detected by the extended source detection algorithm in Section 3.1.

Table 1. Parameters of artificial LV dwarf galaxies.

Parameter	Minimal	Maximal	Step (logscale)
Stellar mass ($M_\odot$)	10^3	10^6	$10^{0.2}$
r_h (pc)	10	316	$10^{0.5}$
D (kpc)	316	19 952	$10^{0.25}$

mock images with comprehensive instrumental and observational features.

To accurately model the impact of the optical system on image quality, a comprehensive simulation model of the CSST optical system has been developed to generate high-fidelity point spread functions (PSFs). This optical simulator consists of six distinct modules that account for various optical aberrations, including mirror surface roughness, fabrication imperfections, CCD assembly errors, and thermal-induced distortions. Additionally, the simulator incorporates two dynamic error sources, micro-vibrations and image stabilization effects, providing a realistic representation of the PSF under operational conditions.

To produce realistic mock images, various noises have been included, such as shot noise, sky background noise, and detector-related effects. Using the throughput system of CSST, photons from each galaxy are generated with Galsim[3] (Rowe et al. 2015). Here, the throughput system accounts for mirror efficiency, filter transmission, and the detector's quantum efficiency, ensuring that the simulated images closely match the actual observational conditions.

Poisson noise is included to model contributions from both the sky background and the CCD detector's dark current. Specifically, the *i*-band background level was set to 0.212 e^-/pixel/s, with a dark current of 0.02 e^-/pixel/s. For a 150s exposure, this results in an average signal of approximately 35 e^-/pixel. Additionally, read noise was modelled as a Gaussian distribution with a standard deviation of approximately 5.0 e^-/pixel. To simulate the generation of mock galaxy images on the detector, bias effects were included, and the gain factor was applied for calibration.

We first generate mock images based on the fiducial catalogue, incorporating basic instrumental effects except for cosmic rays, hot pixels, bad columns and Charge Transfer Inefficiency (CTI). These images, referred to as 'fiducial images'. To facilitate flexible testing of LV dwarf galaxies detection, mock images of LV dwarf galaxies are generated independently. They omit bias, dark current, and sky

[3]https://github.com/GalSim-developers/GalSim

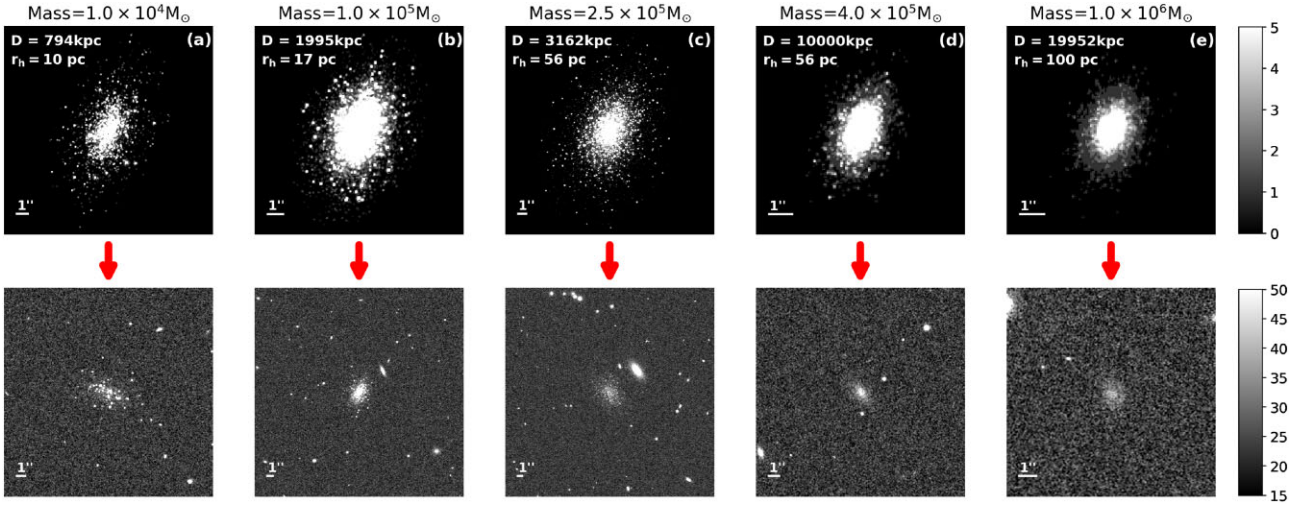

Figure 3. Simulated images of LV dwarf galaxies in the *g* band. The top row shows background-free images of simulated LV dwarf galaxies with a stellar mass of 10^4–10^6 $M_\odot$. The stellar mass, distance, and half-light radius of each simulated galaxy are annotated in the corresponding panel. The bottom row presents the same galaxies as they appear in the fiducial images (background-included).

background. By treating LV dwarf galaxies images as background-free overlays, this enables us to place them freely into the fiducial images. Fig. 3 shows examples of simulated LV dwarf galaxy images in the *g*-band. The first row presents background-free images annotated with key parameters, including stellar mass, distance, and half-light radius. The bottom row displays the same galaxies injected into fiducial images to evaluate detection performance under realistic observational conditions.

3 PRE-PROCESS

LV dwarf galaxy searches start with an automated process of detecting extended sources. This procedure is implemented to the fiducial images that contain artificial LV dwarf galaxy images. Based on the detected sources from the background and from the injected LV dwarf galaxies, we built negative and positive samples to train the AI classifier (see Section 4.1). The full implementation of both algorithmic steps is publicly available.[4] Here, we outline the extended source detection process, which follows a similar approach to that of Bennet et al. (2017), Carlsten et al. (2020), and Davis et al. (2021b).

3.1 Extended source detection

The extended source detection process is applied to the mock images, where high-surface-brightness (HSB) systems are first identified. Masking these bright sources then facilitates the detection of low-surface-brightness (LSB) systems. We divide each mock image (9232 × 9216 pixels) into small areas of 1000 × 1000 pixels, with an overlap of 100 pixels between adjacent areas. The detection procedure is performed in each area as outlined below and also illustrated in Fig. 4:

(i) A background-free image of a simulated LV dwarf galaxy (panel a) is inserted into the fiducial image (panel b). The location of the LV dwarf galaxy is marked with a red circle with thin dashed line.

(ii) The fiducial map (panel b) is convolved with a 3 × 3 pixel Gaussian kernel to produce a smoothed map. This map is then scaled by the local background noise to generate a S/N smoothed map, as shown in panel (c).

(iii) Candidate sources are identified by applying S/N thresholds to define two regions: S_1 with S/N > 6, and S_2 with S/N > 1.5. The sources satisfy all three criteria below are classified as HSB candidates highlighted by the blue circles in panels (c) and (d);

(a) $S_1 > 60$ pixel^2;
(b) $S_2 > 600$ pixel^2;
(c) $S_1 / S_2 > 0.2$.

The sources satisfied only criteria (i) and (iii) are classified as bright source contaminants. These typically correspond to bright stars or compact galaxies that do not exhibit the characteristic profiles expected for LV dwarf galaxy candidates. These contaminants are shown by the white regions without overlaid blue circles with solid lines in panel (d).

(iv) The fiducial map (panel b) is masked by excluding the regions corresponding to the identified HSB sources and bright contaminants (from panel d). The masked regions are filled with local background noise, resulting in a masked map shown in panel (e).

(v) The masked map (panel e) is smoothed again using a 6 × 6 pixel Gaussian kernel and scaled by the background noise to produce a masked S/N smoothed map, as shown in panel (f). LSB candidates are detected in panel (f) by selecting regions where S/N > 4 and $S_2 > 400$ pixel^2. These candidates are highlighted with green circles with dashed lines. The injected LV dwarf galaxy in this example meets the LSB selection criteria.

(vi) Image cut-outs are generated for both HSB and LSB candidates by extracting regions of size 4 × S_2 around each detection, as shown in the rightmost column of figure. For the example LV dwarf galaxy, which meets the LSB criteria, corresponds to the bottom panel in that column. Finally, overlapping detections are cross-matched and merged to eliminate duplicates.

The threshold selection was determined by testing the trade-off using simulated dwarfs with 10^5 $M_\odot$, size 100 pc, and distances of 5 ∼ 10 Mpc as typical targets, corresponding roughly to Local Volume dwarf galaxies which are relatively difficult to

[4] https://github.com/nemoqh77/LVdgdetection/tree/main

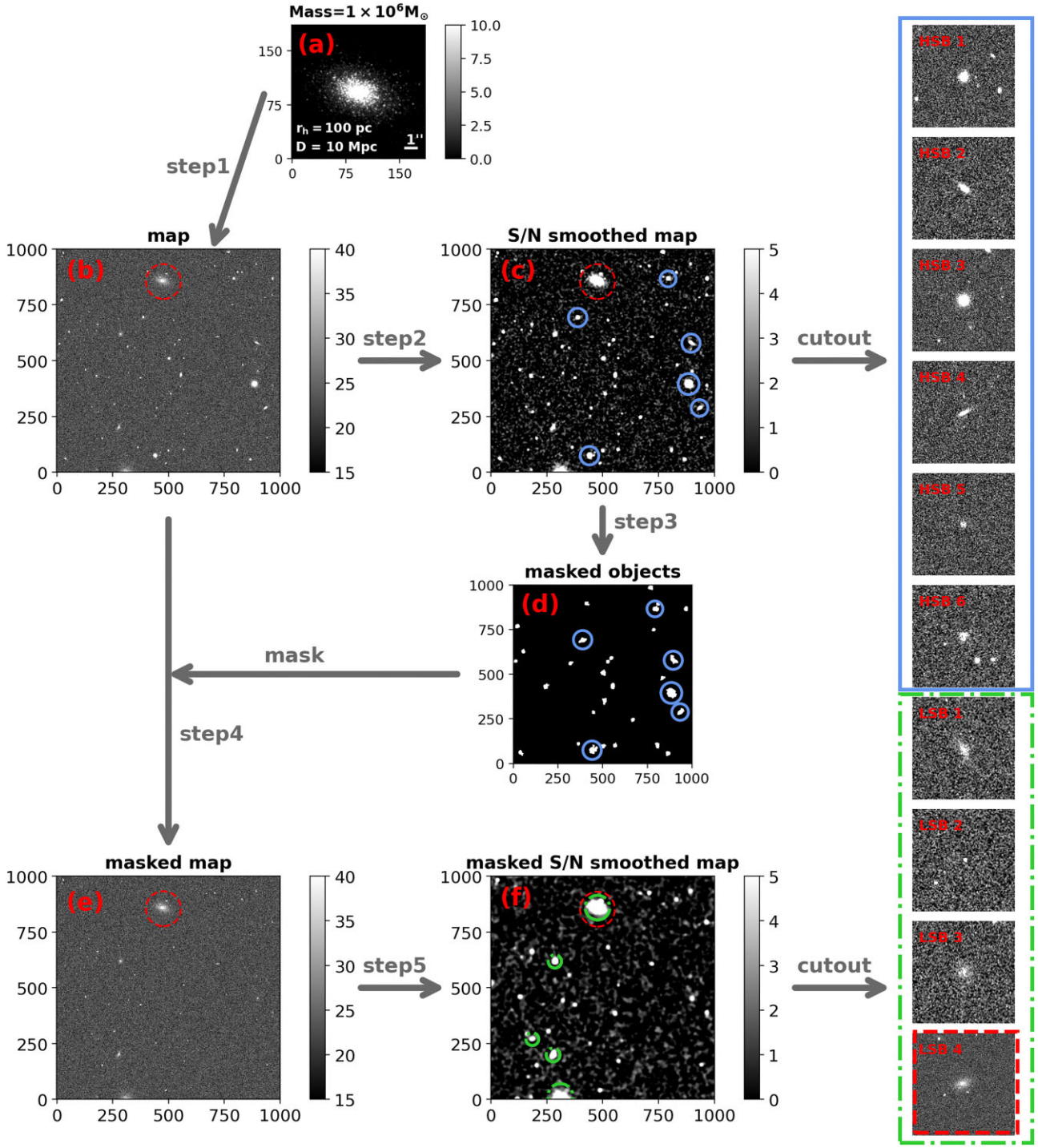

Figure 4. Extended source detection process. (a) Background-free image of a simulated LV dwarf galaxy with a stellar mass of 10^6 $M_\odot$, a half-light radius of 100 pc, and a distance of 10 Mpc. (b) The same galaxy injected into a 'fiducial Image'. (c) S/N map from (b) after convolution with a 3-pixel kernel and scaling by background noise. (d) Sources satisfying the masking threshold; blue circles with solid lines highlight those meeting the HSB candidate criteria. (e) Map (b) after applying mask. (f) Masked S/N map from (e) after convolution with a 6-pixel kernel and scaling by background noise. Green circles with dashed lines highlight those meeting the LSB candidate criteria, the LV dwarf galaxy satisfies the selection in this case. The rightmost column of panels shows cut-outs of extended source candidates that satisfy the detection thresholds. The injected LV dwarf galaxy corresponds to the bottom panel, marked with a red box with dashed line. The red circle with thin dashed line in (b), (c), (e), (f) panels marks the LV dwarf galaxy's location. Colourbars in panels (a), (b), (e) represent flux intensity, and colourbars used in panels (c) and (f) represent the S/N values.

detect in most current observational surveys. We did not exhaustively explore the full parameter space of simulated dwarfs to optimize the thresholds; such detailed optimization can be conducted later using future real observational data. The current threshold choice is sufficient to achieve detection rates acceptable in our tests.

The above procedure is performed independently in the g, r, and i bands. The results from these three bands are then cross-matched. Only detections with a tolerance of less than 1 arcmin are retained as the final detected extended sources.

We identify the extended source candidates from all four sky regions. For all the sources, we find the corresponding sources in the input catalogue by performing cross-match with 1 arcmin radius. This ensures that the detection procedure is finding 'true' systems.

Overall, about 3 per cent of the input sources are identified, the vast majority of which are stars and galaxies brighter than $r = 20$, represented by the dashed histograms in Fig. 2. There are about 300 extended sources detected from a skypatch of $11.'4 \times 11.'4$, yielding around 24 000 sources from the first region, sky0 (see Fig. 1). These detected sources are used to construct the negative samples for the image-classifier. Using the same extended source detection process, detection tests are conducted on 1953 synthetic LV dwarf galaxies. Each galaxy is randomly rotated and tested through 100 independent injections, each time into a random position within a randomly selected fiducial image from the 100 pointings of the sky0 field. The individual detection rates of each LV dwarf galaxy are summarized in Fig. A3. This procedure yields 62 689 cut-outs from 1143 successfully recovered galaxies, which serves as the basis for constructing the positive sample set.

Fig. 5 presents scaled cut-outs of extended source candidates in the g band. The first and second rows of panels correspond to LV dwarf galaxies and other candidates, respectively. The latter are detected in the fiducial images and are primarily distant galaxies. A clear morphological distinction is observed between the two categories: LV dwarf galaxies often display partially resolved structures with discernible outlines of individual member stars, while distant galaxies appear as unresolved sources lacking visible stellar features. These morphological distinctions form the basis for employing machine learning techniques to classify and separate the two types of sources effectively.

Extrapolating the number of sources from a single skypatch to the full CSST survey footprint of 17 500 square degrees, we anticipate detecting over tens of millions of extended sources as contaminants. Given the enormous number of sources, traditional visual inspection to identify LV dwarf galaxies is almost impossible, and thus an automatic image-based image recognition method is necessary for this task.

3.2 Positive and negative samples

To prepare the construction of the positive and negative samples for the image-classifier, we first standardize the cut-outs of extended source candidates to a uniform resolution of 446 × 446 pixels in three channels which correspond to the three bands. The flux values are background-subtracted and linearly scaled for normalization. In order to make the training process more efficient, we zoom into the central 1/4 region of each image, yielding the final image size of 224 × 224.

To construct the positive training data set, we begin with 1143 LV dwarf galaxies that are successfully detected at least once during the extended source detection stage. Each galaxy is tested through 100 synthetic placements followed by detection, as described in the previous subsection, with each successful detection yielding an image cut-out. The total number of cut-outs per LV dwarf galaxy varies according to its detection rate in the extended source detection process. To prevent bright or easily detectable galaxies from dominating the data set, an upper limit of $N_{\max} = 15$ cut-outs per LV dwarf galaxy is applied. Galaxies with fewer than $N_{\max}$ detections contribute all their available cut-outs, ensuring a balanced distribution across systems of varying detectability. For comparison, we also perform experiments with $N_{\max}$ values of 30, 50, 75, and 100 (see Section 4.3 for details).

From these 1143 LV dwarf galaxies, we randomly select 300 galaxies (with their associated cut-outs) as the training set, 200 as the validation set, and assign the remaining 643 galaxies to the testing set. These three subsets are mutually exclusive: each LV dwarf galaxy (along with its associated cut-outs) is assigned to only one group. This ensures that no galaxy appears in more than one subset, so the model's performance on the testing set reflects genuine generalization. We refer to this configuration as Group A. In this setting, over half of the LV dwarf galaxies are assigned to the testing set. Then we construct a corresponding Group B for each Group A configuration. Specifically, from the 643 LV dwarf galaxies in the Group A testing set, we randomly select 300 for training and 200 for validation in Group B. The remaining 143, combined with the 500 LV dwarf galaxies previously used in Group A's training and validation sets, form Group B's testing set. Group A and Group B are used independently for model training and evaluation. This design ensures that all LV dwarf galaxies appear in at least one testing set across Groups A and B, thereby enabling us to assess the ViT model's classification performance for each LV dwarf galaxy.

Negative samples are drawn from extended sources detected in the fiducial images, with ∼30 000 sources in each sky region (as shown in Fig. 1). In sky0, which serves as the primary region for training and validation, a total of 29 152 cut-out images are obtained from the extended source detection pipeline. To ensure class balance, we randomly select 3200 and 2200 samples as negative examples for the training and validation sets, respectively, matching the size of the corresponding positive sets. The remaining 23 752 samples are reserved as the negative component of the testing sets. To further assess model generalization, we construct three additional testing sets using extended source detections from the remaining sky regions (sky1, sky2, and sky3), which contain 27 298, 26 255, and 32 079 cut-outs, respectively.

This sampling strategy is applied independently for each of the 10 Group A/B data set pairs constructed under the $N_{\max} = 15$ configuration, yielding 20 distinct data sets. Table 2 summarizes the sample sizes for Group A. For each data set, the negative training and validation samples are randomly drawn from sky0. The remaining cut-outs from sky0-after removing those selected for training and validation-constitute the testing set for that specific data set. As a result, the exact composition of the negative testing sets differs across the 20 data set pairs.

4 IMAGE CLASSIFICATION

4.1 ViT model

The identification of LV dwarf galaxies can be translated to binary classification task. The model used for this task is the ViT (Dosovitskiy et al. 2020), an advanced architecture based on the Transformer framework. The ViT model divides input images into patches of fixed size, and processes them as tokens through a self-attention mechanism. In contrast to CNNs that rely on convolutional operations to extract local features (Rawat & Wang 2017), the ViT model uses a self-attention mechanism to establish relationships across the entire image, making it highly effective at capturing global contextual information. This advantage is particularly beneficial for tasks like galaxy classification, where understanding spatial and structural relationships is critical. By pre-training on large data sets

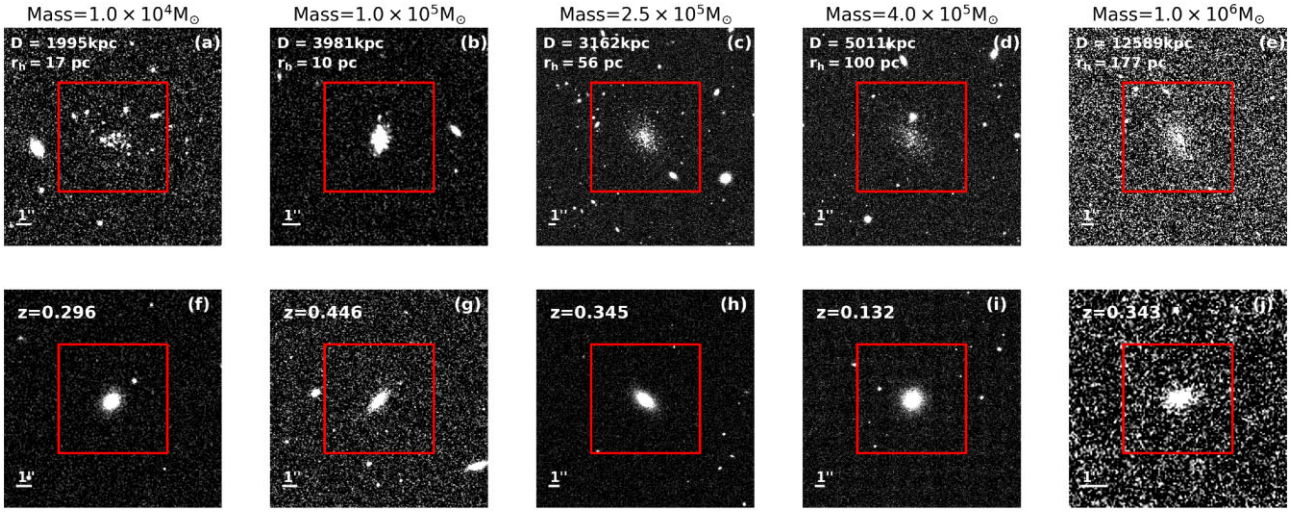

Figure 5. Scaled cut-outs of extended source candidates in the *g* band, obtained from the extended source detection process. The first row shows true LV dwarf galaxies. The second row shows negative samples detected in the fiducial images; these are primarily distant galaxies, with their respective redshifts indicated in the image annotations. The red solid box marks the central quarter of the image, which serves as the input to the ViT model.

Table 2. Group A of ten sets in sky0.

Positive/negative	Training set	Validation set	Testing set
Set 1	3152/3200	2088/2200	6929/23752
Set 2	3290/3200	2136/2200	6743/23752
Set 3	3152/3200	2000/2200	7017/23752
Set 4	3195/3200	2251/2200	6723/23752
Set 5	3176/3200	2118/2200	6875/23752
Set 6	3262/3200	2018/2200	6889/23752
Set 7	3242/3200	2216/2200	6711/23752
Set 8	3274/3200	2129/2200	6766/23752
Set 9	3049/3200	2255/2200	6865/23752
Set 10	3189/3200	2197/2200	6783/23752

such as ImageNet-21k, a data set containing over 21 000 categories of images, the ViT learns transferable features that can be fine-tuned for specific astronomical applications, ensuring robust performance even in the presence of variations in image quality, resolution, and noise.

In our implementation, we train the 'vit-base-patch16-224-in21k' model using the candidate cut-outs obtained from Section 3.2. The ViT variant is pre-trained on the ImageNet-21k data set, which contains 14 million images across 21 000 categories, providing a robust foundation for transfer learning.

4.2 Classification results

We evaluate the classification performance of the ViT model on the testing sets, following training and validation using the data sets described in Section 3.2. We trained ViT models on all 20 data sets use a learning rate of 5×10^{-6}, and the results are summarized in Fig. 6. The false positive rate (FPR) is fixed to 0.001 to ensure a low contamination level. The left panel of Fig. 6 displays the learning curve of the model, showing the true positive rate (TPR) as a function of training epochs. The red solid line denotes the mean TPR, and the shaded region represents the 1σ uncertainty. The model converges to a mean TPR of approximately 85 per cent at around the 60th epoch. The middle panel illustrates the learning curves for different learning rates, evaluated on the Set 1 data set at FPR = 0.01. The middle panel of Fig. 6 presents the learning curves under various learning rates, evaluated on the Set 1 data set at a fixed FPR of 0.01. For comparison, we also include the performance of two CNNs, represented by the grey and pink lines with circular dot markers , which converge to a TPR of approximately 70 per cent. In contrast, the ViT model, shown as the blue and dark red curves, achieves significantly higher TPRs of around 90 per cent, demonstrating superior classification performance under the same FPR constraint. More comparisons of different learning rates are shown in Fig. A2. The right panel shows the TPR–FPR relation after 60 training epochs, with each line representing a different data set. The consistency across data sets highlights the stability and robustness of the ViT model. In addition, based on tests using the Set 1 data, we observe no significant difference in model predictions across the four sky regions (sky0 to sky3) as shown in Fig. A2, further supporting the model's generalization capability.

To further evaluate classification accuracy on a per-galaxy basis, we aggregate the results from all 20 data sets. We define the classification recall of a given LV dwarf galaxy as the TPR at a fixed FPR of 0.001, as illustrated in Fig. 7. Each panel displays the classification recall for LV dwarf galaxies of the same stellar mass, plotted as a function of distance (*x*-axis) and half-light radius (*y*-axis), with colour encoding the classification recall. The results indicate that detection rates are comparable across galaxies with different morphological properties, demonstrating that the ViT model effectively generalizes over key physical properties of LV dwarf galaxies.

4.3 Effects of imbalanced data sets

In supervised learning, the training results depend heavily on how the data sets are constructed. In this work, there are two key parameters in constructing the data sets that would have an impact on the training: the selection of $N_{\rm max}$ and the number of artificial LV dwarf galaxies used to draw training and validation sets (see Section 3.2).

In the baseline configuration ($N_{\rm max} = 15$), we demonstrated that per-galaxy sampling caps effectively mitigate sample imbalance and lead to comparable classification recall across different LV dwarf galaxies (Fig. 7). To further quantify the impact of $N_{\rm max}$, we explored

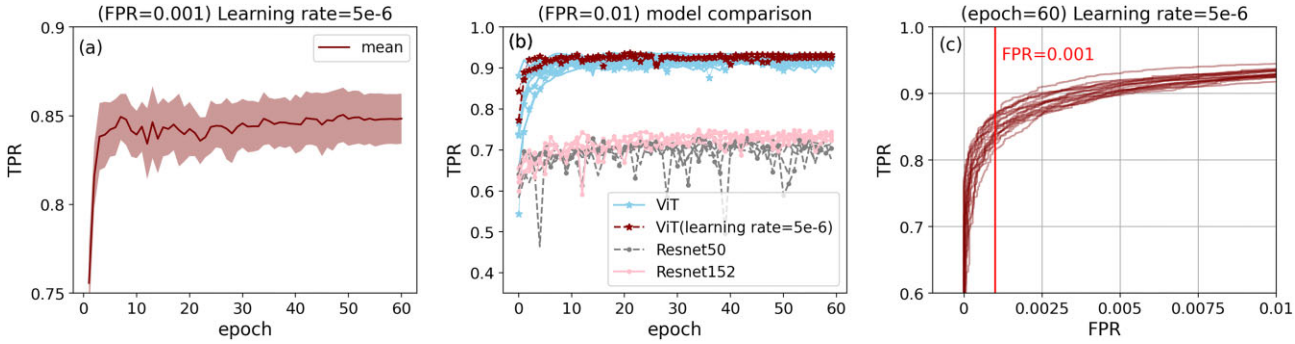

Figure 6. Performance evaluation of different models and training configurations based on true positive rate (TPR). Panel (a): Evolution of TPR during training at FPR = 0.001. The red solid line indicates the mean TPR across 20 data set samples, and the shaded region denotes the 1σ deviation. Panel (b): Comparison of training performance among ResNet50, ResNet152, and ViT models, trained on linearly scaled sky0 data using $N_{\max} = 15$ (Set 1, Group A). The x-axis represents training epochs, and the y-axis shows the TPR at FPR = 0.01. Pink, grey, and blue lines with distinct markers correspond to ResNet50, ResNet152, and ViT, respectively. Panel (c): Classification performance of ViT at epoch 60. Each curve shows the TPR–FPR relation for one of 20 independent testing set samples.

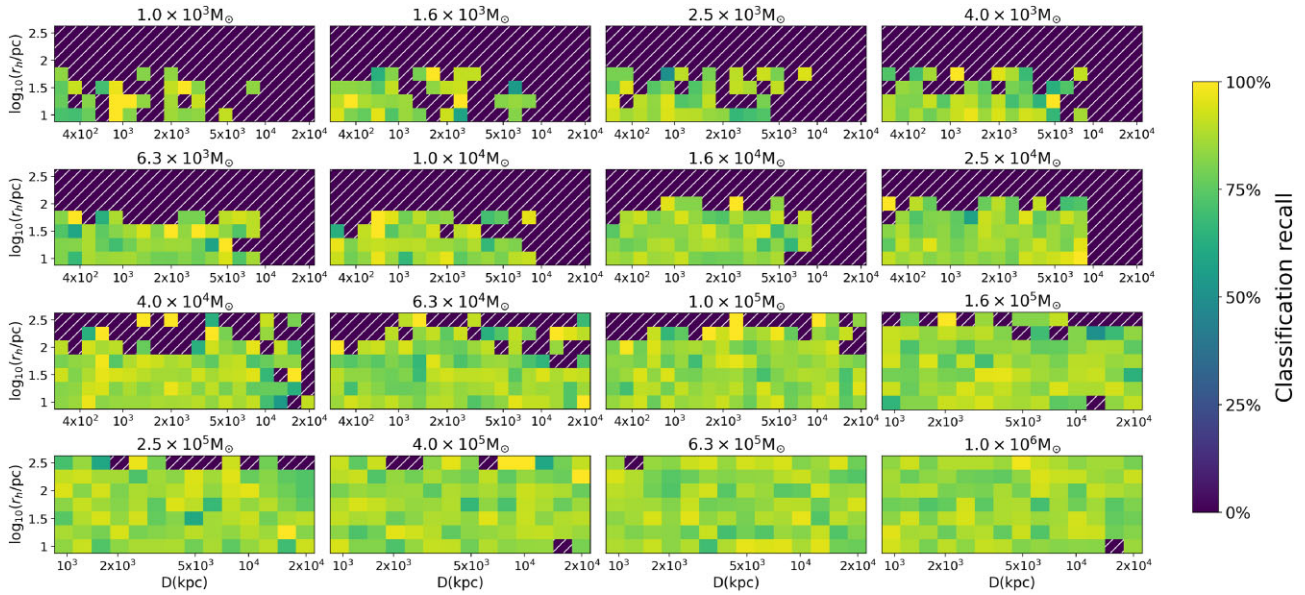

Figure 7. Mean classification recall of the ViT model for LV dwarf galaxies across all 20 data sets. The colour of each pixel corresponds to the TPR at FPR = 0.001 for the respective LV dwarf galaxies. The shaded regions indicate areas without test samples.

larger caps. As shown in Fig. 8, the completeness steadily increases with $N_{\max}$, reaching TPR $\gtrsim 90$ per cent for $N_{\max} \geq 30$. However, a closer look reveals that this gain is driven primarily by brighter, nearby galaxies, while fainter systems suffer reduced classification recall when high-detection galaxies dominate the training set (Fig. 9, pixels with red dots).

To mitigate this imbalance, we generated an augmented data set by increasing the number of detection tests for galaxies with pre-process detection rates below 30 per cent, adding 200 additional cut-outs per source. This balanced configuration, denoted as $N_{\max} = 30$ (balance), yields performance comparable to the baseline $N_{\max} = 15$ set-up (green line with triangle markers in right panel of Fig. 8), with a TPR of ∼85 per cent at FPR = 0.001. These results highlight that, although larger $N_{\max}$ values can improve completeness, careful balancing is required to maintain sensitivity to faint systems and avoid bias toward bright galaxies.

The sampling strategy is another key factor influencing model performance. In the fiducial configuration, we draw ∼3200 cut-outs from 300 LV dwarf galaxies for training, ∼2200 cut-outs from 200 galaxies for validation, and assign the remaining 643 galaxies to the testing set (see Section 3.2). This set-up allocates more than half of the LV dwarf galaxies to evaluation, providing a stringent test of the ViT model's classification ability across a wide variety of galaxies, assuming that morphological similarities within the parameter space of dwarf galaxies allow for effective generalization by the model. However, such a strict configuration limits the training diversity available to the model. To investigate the effect of a more training-heavy sample split, we construct a control set-up, referred to as '$N_{\max} = 15$ (large)', in which the training and validation sets are expanded to include 600 and 400 LV dwarf galaxies, respectively, leaving 143 for testing. As shown in the right panel of Fig. 8, this set-up leads to a TPR exceeding 87 per cent on the testing set (brown–yellow line with circular dot markers). The result confirms that increasing the number of training examples improves the model's ability to generalize, though at the cost of a smaller testing set.

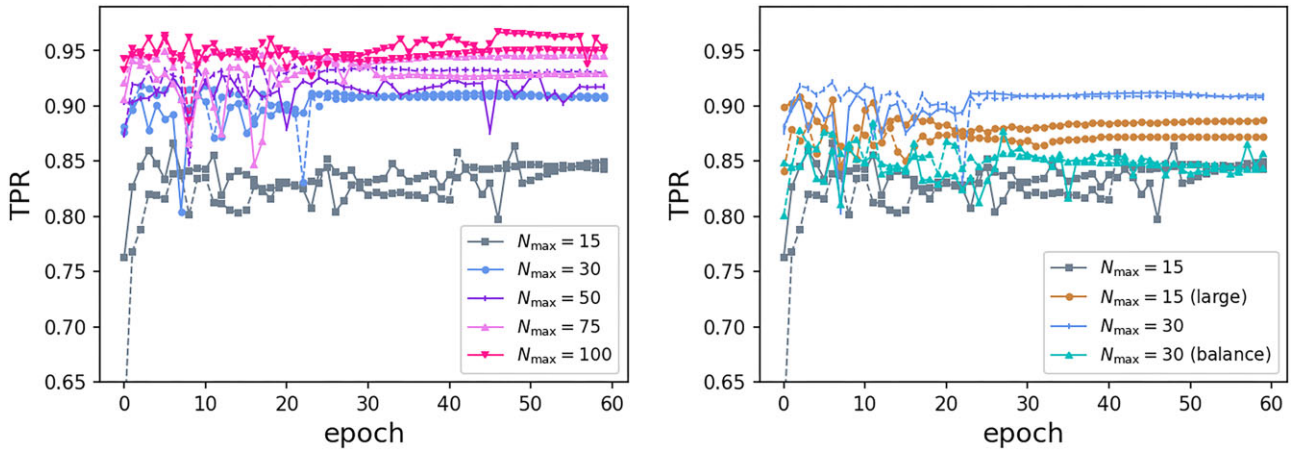

Figure 8. ViT classification recall under different data sets.The x-axis represents the number of training epochs, and the y-axis corresponds to the TPR at FPR = 0.001. Lines in different colours with distinct markers represent results for different N_{max} values or data sets. For lines of the same colour, solid and dashed lines correspond to the results of samples A and B in Set 1.

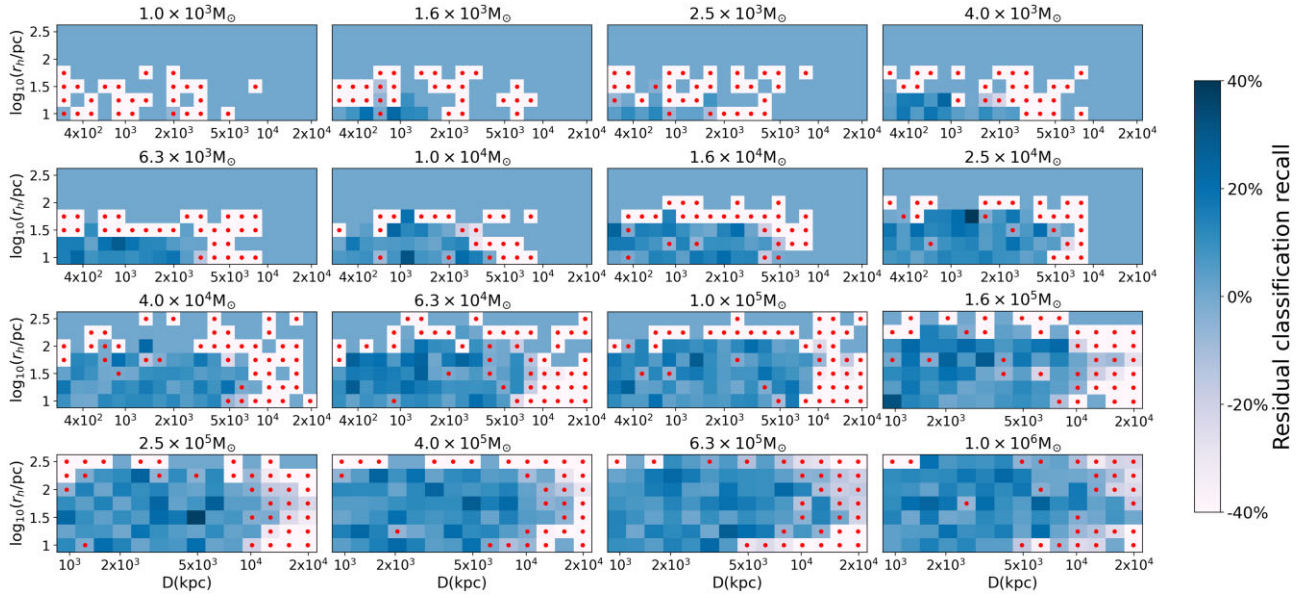

Figure 9. Difference map of ViT classification recall between $N_{max} = 30$ and $N_{max} = 15$ across 20 data sets. The figure shows the difference in TPR distributions (at FPR = 0.001) between $N_{max} = 30$ and $N_{max} = 15$. Pixels with red dots indicate LV dwarf galaxies for which the TPR in the $N_{max} = 30$ group is lower than that in the $N_{max} = 15$ group.

Building on the previous results, we find that increasing the diversity and quantity of LV dwarf galaxies in the training set improves classification performance. However, excessively large training sets, especially those derived from synthetic data may lead to overfitting, potentially limiting model generalizability when applied to real observations. Because our current data set is based on mock images, a more rigorous evaluation will require testing against real observational data. While archival images of known LV dwarf galaxies are available from existing surveys, their quality and characteristics are highly dependent on the real instrumentation performance, particularly the optical resolution. A more robust assessment will thus only become feasible once CSST survey data become available.

The primary goal of this work is to establish and validate a complete detection pipeline using controlled simulations. Once CSST observations commence, the model will need to handle additional complexities such as crowded stellar fields, proximity to large galaxies, and background contamination from diffuse light. These environments can significantly affect detection reliability. Future adaptations may also involve tuning extended source detection thresholds and retraining on hybrid data sets combining mock and real observations to enhance robustness.

5 POST PROCESSING

In constructing the ViT model, we noticed that some LV dwarf galaxies are misclassified, as shown in panel (a) in Fig. 10 (more examples are shown in Fig. A1). A common feature observed in some of these galaxies is the presence of resolved member stars and a central concentration of stellar components. These features allow us to distinguish them from background galaxies (as shown in the second row of Fig. 5), which generally lack such central overden-

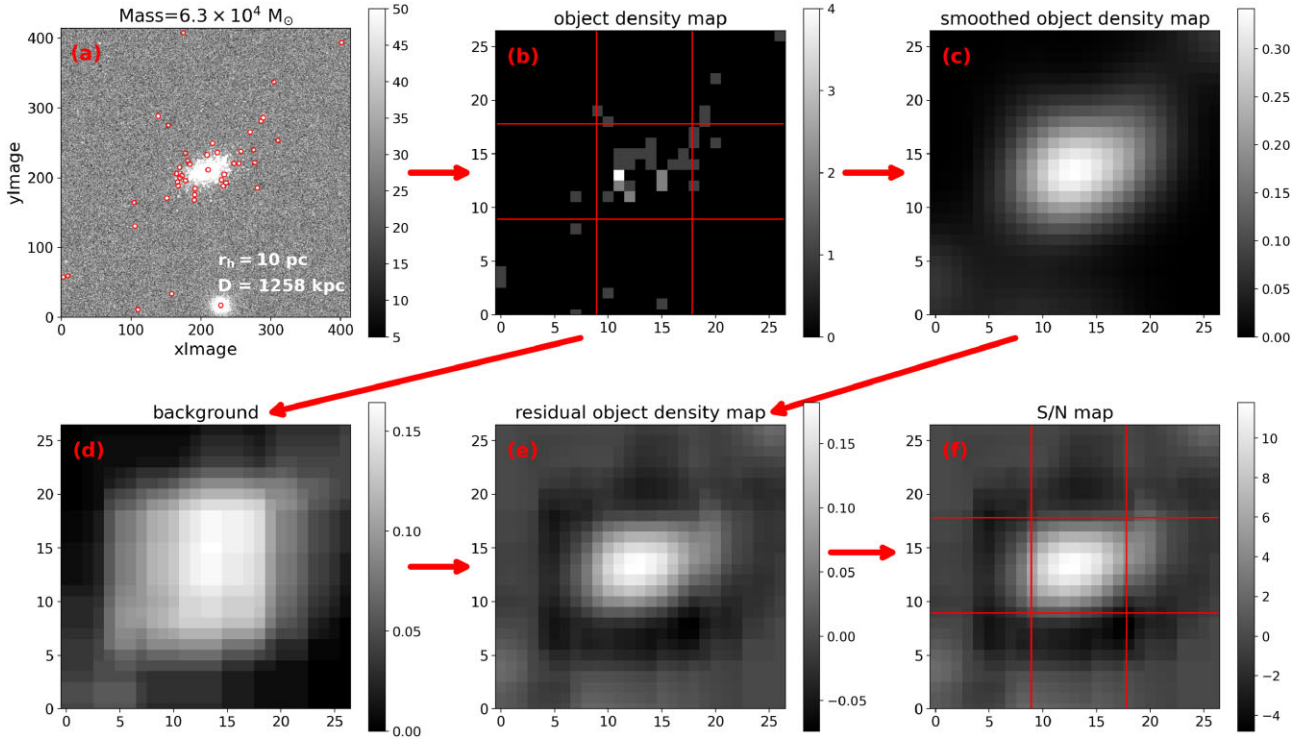

Figure 10. Post-Process Workflow Illustration. (a) Cut-out image with red circles with solid lines marking the objects detected by the Source Extractor. (b) Density map of objects corresponding to (a), with each pixel representing 1 arcsec. (c) and (d) are the results of convolving (b) with kernels of 'σ = 1' and 'σ = 28,' respectively. (e) Residual density map obtained by subtracting (d) from (c). (f) S/N map derived by dividing (e) by the background noise.

sities. Motivated by this morphological distinction, we introduce a post-processing step to enhance the classification accuracy of LV dwarf galaxies.

In the post-processing, we re-examine all samples initially classified as negative by the ViT model, evaluating their central overdensities based on both the number of objects and their spatial concentration in the central region. This additional step allows us to identify potential LV dwarf galaxy candidates that are missed by the ViT classifier. As a result, the recovery rate of true LV dwarf galaxies can be improved without significantly increasing the FPR.

5.1 Central overdensity evaluation

The post-processing is performed on unscaled cut-out images, with dimensions corresponding to the full region displayed in panel (h) of Fig. 4. For each sample, we use Source Extractor to identify sources in both the g- and i-band images, and retain only the matched detections, hereafter referred to as 'SE-detections'. We then assess potential source overdensities at the centre of each cut-out.

The detailed procedure is illustrated in Fig. 10. Panel (a) displays the spatial distribution of SE-detections in the cut-out image. Red circles with solid lines indicating 'SE-detections' identified in both the g- and i-band images. The cut-out region is divided into 1 arcsec/pixel bins, and the number of SE-detections in each bin is used to construct the object density map (panel b). To enhance potential overdensity signals, this map is convolved with a Gaussian kernel of width σ_1 (panel c), while a broader Gaussian kernel (σ_2) is applied to generate a background reference (panel d). Subtracting the background from the smoothed density map yields the residual object density map (panel e). The S/N map is then obtained by normalizing the residual map with the standard deviation of the background, estimated from the outer two-thirds of the region (panel f). Finally, we measure two key quantities within the central third of each cut-out: the central object count (from panel b) and the centre S/N (from panel f). These metrics serve as indicators of central source overdensities and are used in the subsequent classification refinement.

In Fig. 11, we show the distributions of 'central object count' and 'central S/N' for all samples initially classified as negative by the ViT model, including both true positives (actual LV dwarf galaxy) and true negatives. The vertical axis indicates the percentage of samples. The results demonstrate that LV dwarf galaxy samples exhibit significantly higher central overdensities than true negatives. Based on this distinction, we apply the following selection criteria: samples with 'central object count' > 7 and 'central S/N' > 6 are reclassified as LV dwarf galaxies.

After applying the post-processing step, approximately 45 per cent of LV dwarf galaxies initially misclassified as negative by the ViT model are correctly reclassified as new positive. Meanwhile, the contamination rate (the proportion of true negatives incorrectly reclassified as LV dwarf galaxies) increases by only 0.02 per cent.

Fig. 12 shows the post-process detection rate distribution for LV dwarf galaxies. A clear trend emerges: brighter, nearer, and more spatially extended systems are more likely to be recovered. This is attributed to their higher number of resolvable member stars and denser spatial profiles, making them more distinguishable. These findings confirm that the post-process step significantly improves LV dwarf galaxy classification accuracy with minimal false positives. This serves as a complementary approach to the ViT-based classifier, which would optimize the detectability across different distances. By incorporating the post-process results into the ViT classification

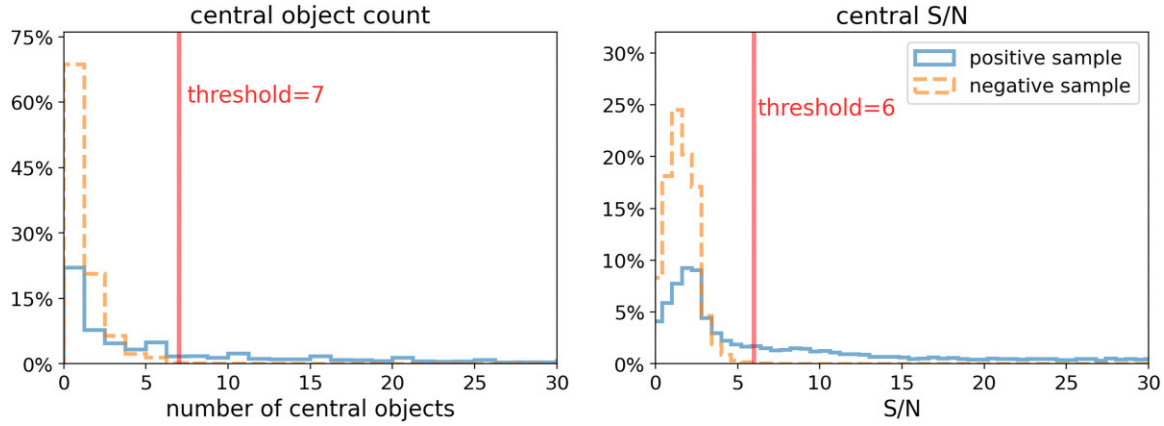

Figure 11. Distribution of 'central object count' and 'central S/N' for all samples classified as negative by the ViT model, including true positive samples (blue solid line) and true negative samples (orange dashed line). The vertical axis of the histogram represents the percentage of samples. Red lines indicate the thresholds used for LV dwarf galaxy selection in the post-process step.

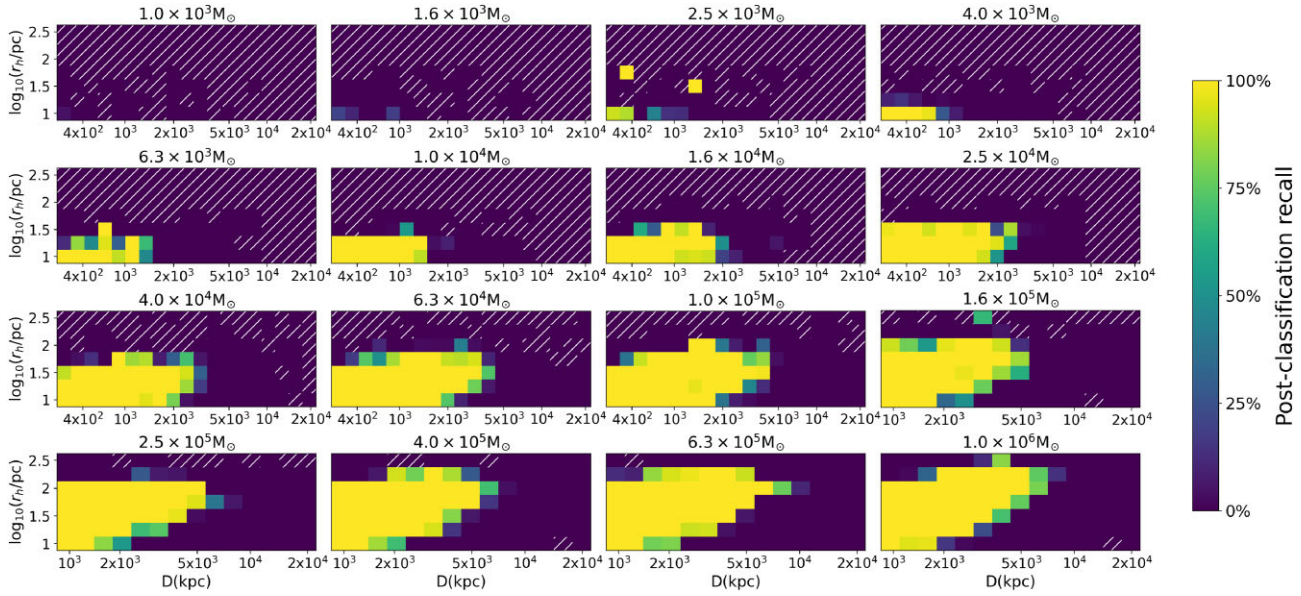

Figure 12. Classification recall during the post-processing stage. The shaded regions indicate areas without test samples.

output, the success rate of LV dwarf galaxy identification is further improved beyond what is shown in Fig. 7. This enhancement is illustrated in Fig. A4.

5.2 Overall detection rate

As previously outlined, our pipeline consists of three main steps: pre-processing, ViT classifier, and post-processing. By combining the detection and classification completeness achieved at each of these stages, we derive the overall detection efficiency of our LV Dwarf Galaxy Detection Pipeline, as shown in Fig. 13. This figure provides a comprehensive assessment of this work in identifying and classifying dwarf galaxies within the local volume.

Fig. 14 compares our detection results with those of currently known nearby dwarf galaxies. To represent the observational sample, we incorporate two Local Volume galaxy catalogues: LVG-large (Karachentsev et al. 2013) and LVGDB-2024 (Pace 2024). The LVGDB-2024 catalogue (marked by red circles in Fig. 14) serves as our primary reference, as it includes all known dwarf galaxies within 3 Mpc and reliable photometric measurements. The LVG-large catalogue (indicated by pink plus signs) is used as a complementary data set. Due to its broader coverage, we only include galaxies beyond 1 Mpc that are not already present in LVGDB-2024. Since LVG-large does not provide direct values for M_V or surface brightness, we estimate them based on a_{26} (the semimajor axis at the 26 mag/arcsec^2 isophote) and m_{26} (the integrated magnitude within that isophote), which may introduce small deviations from true values.

In our previous study (Qu23), we employed a classic matched-filter technique to evaluate CSST's detection capabilities for Local Group dwarf galaxies, using only the simulated stellar catalogues. In the present work, we build upon that analysis by comparing the detection limits derived from both methods within overlapping parameter spaces. The green solid lines and blue dashed lines in Fig. 14 represent the detection limits obtained in this work and in Qu23, respectively. Compared to existing observations, our detection limits reach fainter magnitudes at fixed distances within 20 Mpc, demonstrating the advantage of the image-based method proposed in this work for detecting smaller and more distant dwarf galaxies.

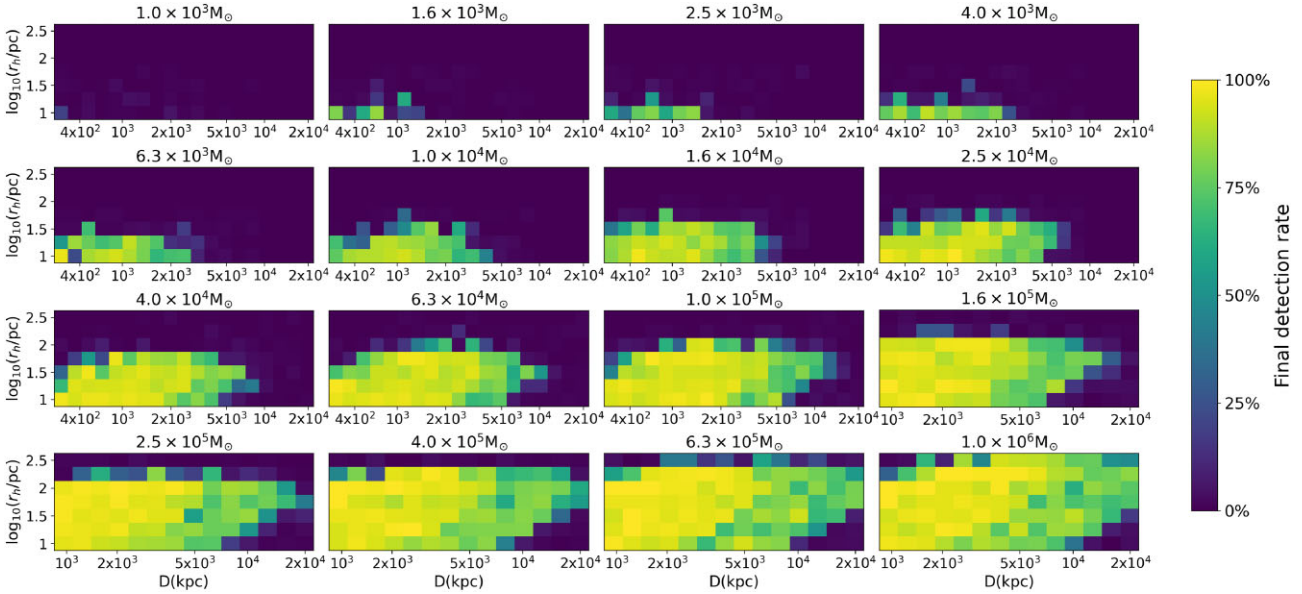

Figure 13. Combined detection rate for LV dwarf galaxies, obtained by combining the ViT prediction results with the post-processing step (Fig. A4) and multiplying them with the detection rates from Step 1 (Fig. A3).

However, beyond 3 Mpc, the surface brightness of known dwarf galaxies typically falls below our detection threshold. Some of these systems have been primarily discovered through small-area, deep-exposure surveys targeting satellite populations, often focusing on specific host galaxies. Such targeted observations allow for longer exposure times and thus reach deeper magnitude limits than those achievable by CSST. In contrast, the main strength of CSST lies in its significantly larger sky coverage, enabling a more uniform and wide-area search for faint dwarf systems. Moreover, certain specialized algorithms designed for detecting low surface brightness sources have allowed some studies to reach deeper detection limits, despite using data with limiting magnitudes shallower than those of CSST (Jones et al. 2024). This highlights a current limitation in our work: the pre-processing threshold criteria we adopt are not optimized for identifying low surface brightness galaxies, which is an important aspect we aim to improve in future studies.

Compared to this work, Qu23 achieves higher recovery rates for systems within 500 kpc (owing to their larger angular sizes and more easily resolved stars) but it is less effective for distant or unresolved systems. Although it reaches a fainter surface brightness limit overall, the method proposed in this work is more sensitive in terms of M_V. Notably, beyond 500 kpc, our detection limit in M_V surpasses that of Qu23 by more than one magnitude. None the less, Qu23 remains superior in terms of surface brightness sensitivity within 5 Mpc. For galaxies at distances exceeding 5 Mpc, the absence of resolved bright stars further limits the applicability of the Qu23 algorithm. In contrast, our current approach remains effective in this regime, achieving higher detection rates and showing better suitability for detecting unresolved systems at greater distances.

We note the existence of a dwarf galaxies that lie well below our detection threshold: NGC55-dw1, a satellite of NGC 0055 (McNanna et al. 2024) ($M_V = -8$, distance $= 2.2$ Mpc, $\mu = 32.25$ mag/arcsec^2), which was identified in DES Year 6 data, using an improved matched-filter algorithm that focused on the 300 kpc–2 Mpc range. That study introduced refined colour filtering and assessed local stellar overdensities across multiple scales, achieving a significantly deeper detection threshold. The detection strategy employed in that work (particularly the methodological refinements) offers useful insights for future CSST-based searches.

6 SUMMARY

The search and identification of dwarf galaxies in the Local Volume are crucial for constructing the satellite galaxy luminosity function in nearby systems. To carry out a comprehensive and effective search, the depth and quality of observational data, as well as the efficiency of the detection algorithm, are all essential. The upcoming CSST sky survey provides a new opportunity for the comprehensive search for nearby dwarf galaxies. In this study, we systematically evaluate the detection capabilities of CSST for dwarf galaxies within the Local Volume.

Using the CSST Image Simulator, we generated multiband synthetic images based on the primary survey parameters of CSST, along with a set of mock images representing LV dwarf galaxies spanning a wide range of distances, magnitudes, and structural properties. We developed a three-step detection pipeline consisting of extended source detection from images, and classification using a ViT model to identify LV dwarf galaxies. For dwarf galaxy systems misclassified by the ViT model, the developed post-processing steps can recover around half of them. Within this framework, we quantified the detection and classification completeness for dwarf galaxies in the CSST Wide Survey.

The classification component in our pipeline employs the 'vit-base-patch16-224-in21k' model, which has demonstrated strong performance in identifying nearby dwarf galaxies from simulated CSST imaging data in this work, achieving a TPR exceeding 85 per cent at a fixed FPR of 0.1 per cent. To further improve completeness, a post-process step is introduced to re-examine initially negatively misclassified dwarf systems, enabling the recovery of originally missed LV dwarf galaxy candidates. This step increases the overall TPR to approximately 92 per cent, without a significant increase in FPR.

In comparison with the method proposed in Qu23, we find that the two algorithms exhibit complementary detection capabilities across

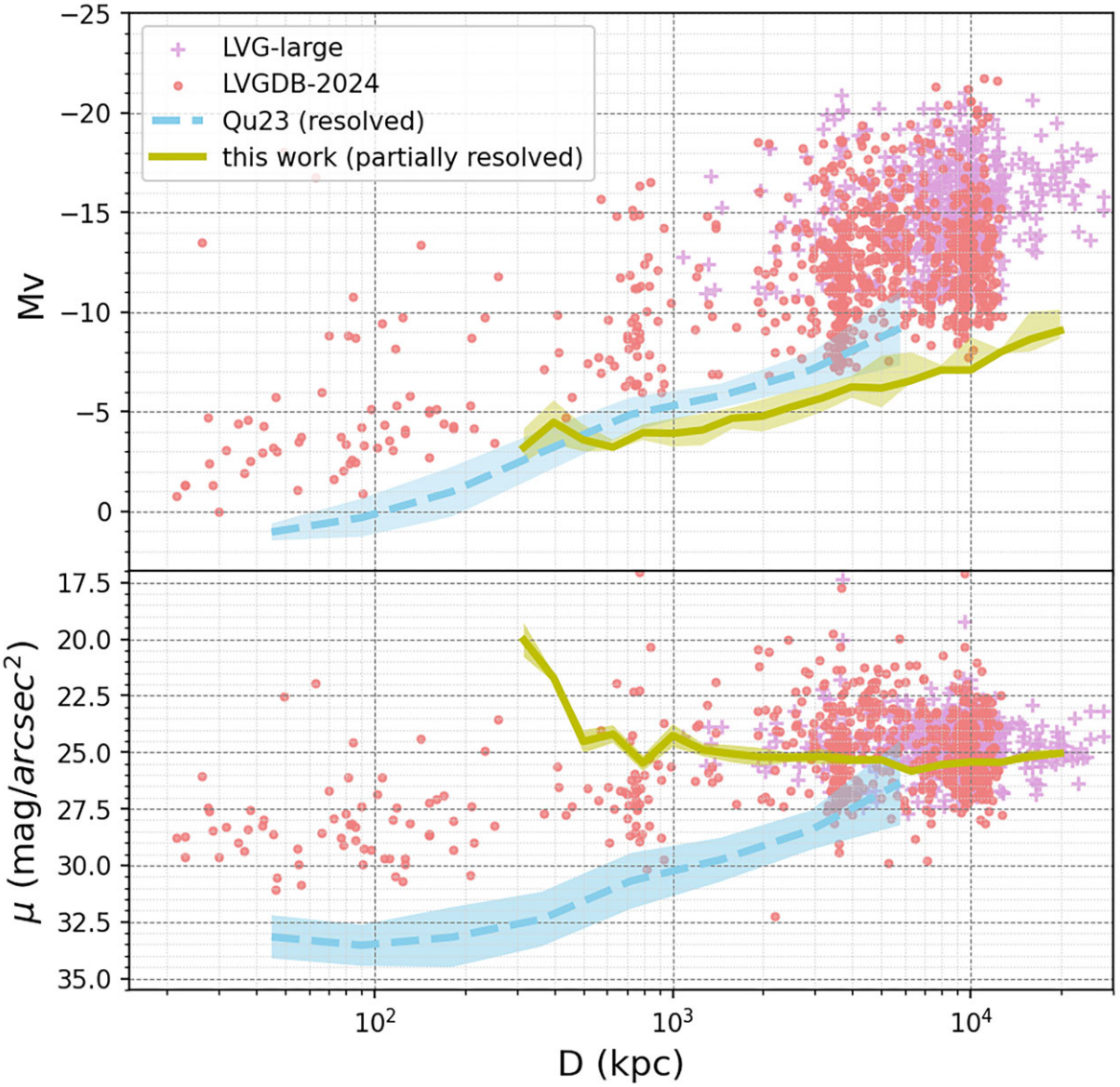

Figure 14. Comparison between the CSST dwarf galaxy detection limits derived in this work and the distribution of observed nearby galaxies. Top panel: Absolute magnitude as a function of distance. Bottom panel: Surface brightness versus distance. Purple plus signs and red circles indicate known Local Volume galaxies (Karachentsev, Makarov & Kaisina 2013; Pace 2024). The green line shows the detection limits from this study, with the solid line marking the 50 per cent completeness level and the shaded region spanning the 25 per cent to 75 per cent completeness range. The blue dashed line indicates the detection limit from Qu23 for comparison.

the parameter space defined by galaxy distance, half-light radius, and absolute magnitude. While Qu23 performs more effectively for nearby galaxies with larger angular sizes as well as more easily resolvabled stars, the approach developed in this work is better suited for detecting distant or incompletely resolved systems. These differences underscore the value of combining both techniques to achieve a more complete census of dwarf galaxies in the Local Volume (Medoff et al. 2025). Based on the detection efficiency derived from our pipeline, we find that a 50 per cent detection rate corresponds to a limiting absolute magnitude of $M_V \lesssim -7$ within 10 Mpc, with a surface brightness threshold of $\sim$25 mag/arcsec^2 for dwarf galaxies at 2–5 Mpc, and $\sim$26 mag/arcsec^2 for dwarf galaxies at 5–10 Mpc. These results indicate that our method is well suited for enabling a systematic search for ultradiffuse galaxies across the Local Volume.

We emphasize that the detection algorithm proposed in this study is fully image-based and does not rely on any higher level data products. The only preprocessing steps required are reference image calibration and image stacking. This allows the method to be applied directly to single-exposure pointings. This enables real-time detection and candidate identification during the early stages of the CSST survey. In contrast, the approach presented in Qu23 depends on pre-processed stellar catalogues, underscoring the flexibility and operational independence of our pipeline.

In practice, real observational data present additional complexities not captured in simulations, include higher densities of large background galaxies, increased noise from instrumental and atmospheric effects, and diffuse light contamination from nearby host galaxies.

These factors may affect the completeness and reliability of source detection and classification. To address this, our pipeline incorporates several tunable parameters such as the S/N and size thresholds for extended source detection. The strategies for constructing the training data set can be further optimized once real CSST data become available.

Although our method is designed to be fully independent of external data products, integrating it with supplementary information such as star–galaxy classification, photometric redshifts, or multiband photometry may substantially enhance detection performance. This hybrid approach could be especially valuable for identifying marginal or ultradiffuse systems, and represents a promising direction for maximizing the scientific return of CSST in the study of low surface brightness galaxies.

ACKNOWLEDGEMENTS

We acknowledge the science research grants from the China Manned Space Project with no. CMS-CSST-2025-A11 and CMS-CSST-2025-A20, the cosmology simulation data base (CSD) in the National Basic Science Data Center (NBSDC) and its funding NBSDC-DB-10 (no. 2020000088). We utilized the high-performance computing cluster from the Purple Mountain Observatory of the Chinese Academy of Sciences, which is equipped with GPU acceleration cards made in China.

XK acknowledge the support from the National Key Research and Development Programme of China (no. 2022YFA1602903), the China Manned Space project with no. CMS-CSST-2025-A10.

DATA AVAILABILITY

The code used in this work is available at https://github.com/nemoqh77/LVdgdetection/tree/main.

REFERENCES

Bennet P., Sand D. J., Crnojević D., Spekkens K., Zaritsky D., Karunakaran A., 2017, ApJ, 850, 109
Bennet P., Sand D. J., Crnojević D., Spekkens K., Karunakaran A., Zaritsky D., Mutlu-Pakdil B., 2020, ApJ, 893, L9
Bhavanam S. R., Channappayya S. S., P. K S., Desai S., 2024, Ap&SS, 369, 92
Bressan A., Marigo P., Girardi L., Salasnich B., Dal Cero C., Rubele S., Nanni A., 2012, MNRAS, 427, 127
Carlsten S. G., Greco J. P., Beaton R. L., Greene J. E., 2020, ApJ, 891, 144
Carlsten S. G., Greene J. E., Beaton R. L., Danieli S., Greco J. P., 2022, ApJ, 933, 47
Crosby E., Jerjen H., Müller O., Pawlowski M., Mateo M., Dirnberger M., 2023, MNRAS, 521, 4009
Cuillandre J.-C. et al., 2025, A&A, 697, A6
Danieli S., van Dokkum P., Conroy C., 2018, ApJ, 856, 69
Davis A. B. et al., 2021a, MNRAS, 500, 3854
Davis A. B. et al., 2021b, MNRAS, 500, 3854
Dekker A., Ando S., Correa C. A., Ng K. C. Y., 2022, Phys. Rev. D, 106, 123026
Doliva-Dolinsky A. et al., 2023, ApJ, 933, 135
Doliva-Dolinsky A., Collins M. L. M., Martin N. F., 2026, Encyclopedia of Astrophysics, VOL. 4, p 61 Elsevier, p.
Dosovitskiy A. et al., 2020, preprint (arXiv:2010.11929)
Drlica-Wagner A. et al., 2015, ApJ, 813, 109
Drlica-Wagner A. et al., 2021, ApJS, 256, 2
Engler C. et al., 2021, MNRAS, 507, 4211
Euclid Collaboration, 2022, A&A, 662, A112
Fernández-Iglesias J., Buitrago F., Sahelices B., 2024, A&A, 683, A145
Forouhar Moreno V. J., Benítez-Llambay A., Cole S., Frenk C., 2022, MNRAS, 517, 5627
Girardi L., 2016, Astron. Nachr., 337, 871
Gondhalekar Y., Moriwaki K., 2024, preprint (arXiv:2411.14392)
Governato F. et al., 2015, MNRAS, 448, 792
Gozman K. et al., 2024, ApJ, 977, 179
Han J. et al. 2025, Sci. China Phys. Mech. Astron., 68, 10, 1
Homma D. et al., 2024, PASJ, 76, 733
Huang K.-W., Chih-Fan Chen G., Chang P.-W., Lin S.-C., Hsu C.-J., Thengane V., Yao-Yu Lin J., 2022, Lecture Notes in Computer Science (LNCS), 13685, 143
Ivezić Ž. et al., 2019, ApJ, 873, 111
Jones M. G. et al., 2024, ApJ, 971, L37
Kanehisa K. J., Pawlowski M. S., Heesters N., Müller O., 2024, A&A, 686, A280
Karachentsev I. D., Makarov D. I., Kaisina E. I., 2013, AJ, 145, 101
Koposov S. et al., 2008, ApJ, 686, 279
Laevens B. P. M. et al., 2015, ApJ, 802, L18
Lee J. C. et al., 2008, in Funes J. G., Corsini E. M., eds, ASP Conf. Ser. Vol. 396, Formation and Evolution of Galaxy Disks. Astron. Soc. Pac., San Francisco, p. 151
Lin J., Liao S.-M., Huang H.-J., Kuo W.-T., Hsuan-Min Ou O., 2021, preprint (arXiv:2110.01024)
Martin N. F., Ibata R. A., McConnachie A. W., Mackey A. D., Ferguson A. M. N., Irwin M. J., Lewis G. F., Fardal M. A., 2013, ApJ, 776, 80
Martin N. F. et al., 2016, ApJ, 833, 167
McNanna M. et al., 2024, ApJ, 961, 126
Medoff J. et al., 2025, preprint (arXiv:2504.18645)
Merritt A., van Dokkum P., Abraham R., Zhang J., 2016, ApJ, 830, 62
Pace A. B., 2024, preprint (arXiv:2411.07424)
Planck Collaboration VI, 2020, A&A, 641, A6
Qu H. et al., 2023, MNRAS, 523, 876
Rawat W., Wang Z., 2017, Neural Comput., 29, 2352
Robertson B. E. et al., 2023, ApJ, 942, L42
Rowe B. T. P. et al., 2015, Astron. Comput., 10, 121
Shu Y., Cañameras R., Schuldt S., Suyu S. H., Taubenberger S., Inoue K. T., Jaelani A. T., 2022, A&A, 662, A4
Simon J. D., 2019, ARA&A, 57, 375
Tanoglidis D. et al., 2021, ApJS, 252, 18
Tollerud E., Hamanowicz A., Mao Y.-Y., Geha M., Wechsler R., Weiner B., Nadler E., Kallivayalil N., 2022, American Astronomical Society Meeting #240. p. 145.11
Walsh S. M., Willman B., Jerjen H., 2009, AJ, 137, 450
Zaritsky D. et al., 2019, ApJS, 240, 1
Zhan H., 2021, Chinese Sci. Bull., 66, 1290

APPENDIX A: ADDITIONAL FIGURES

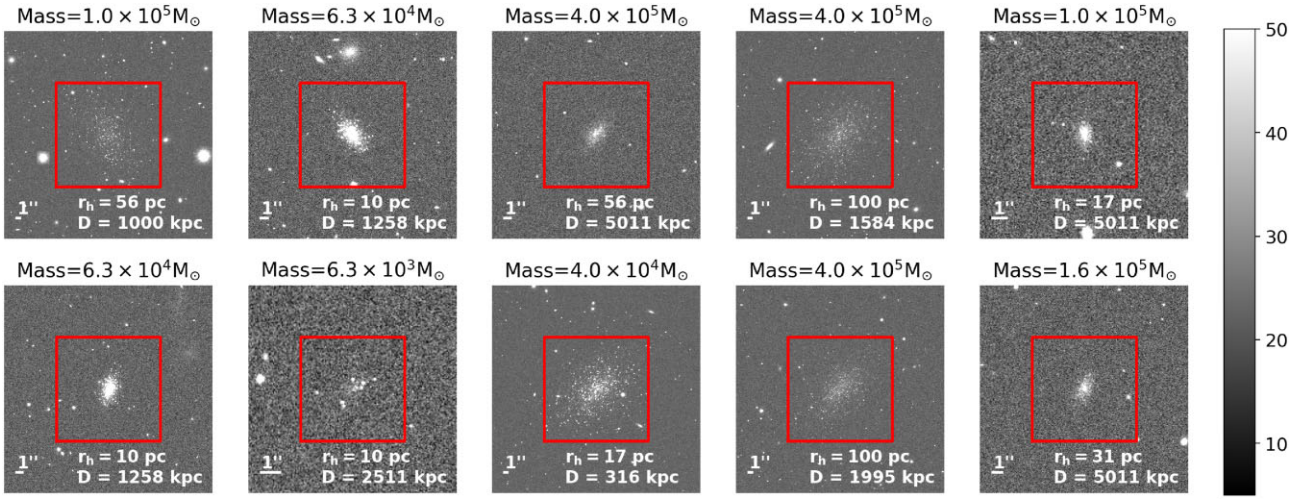

Figure A1. Unscaled cut-outs (in *g* band) of LV dwarf galaxies misclassified by the ViT model.

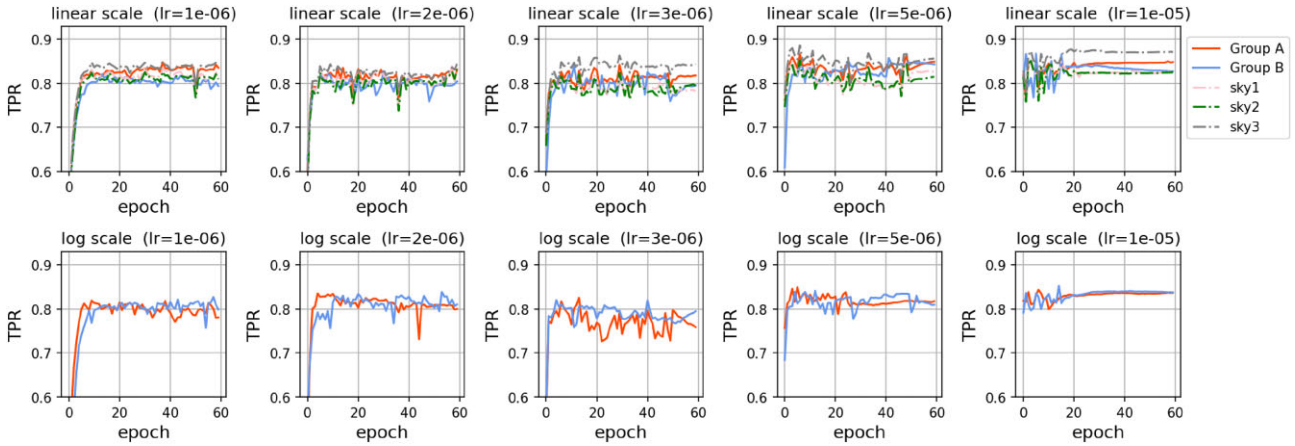

Figure A2. The variation of TPR at FPR = 0.001 across different testing sets with training epochs under various learning rates. The first row presents the results for cut images processed with linear scaling, while the second row shows the results for cut images processed with logarithmic scaling. Each column corresponds to a different learning rate.

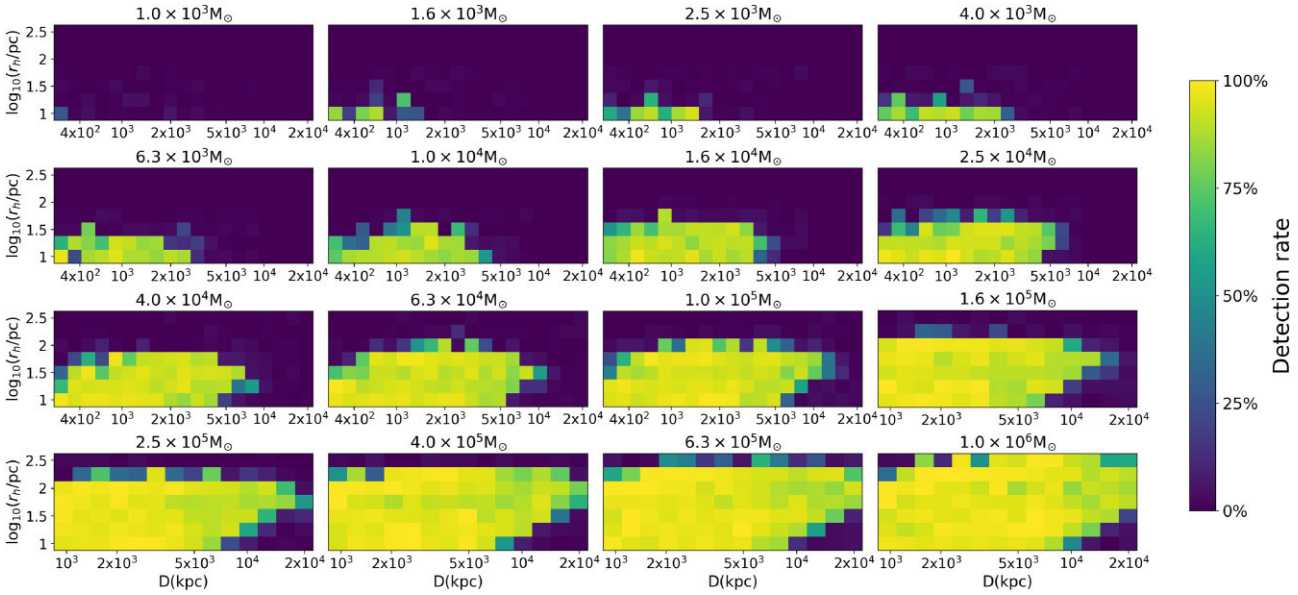

Figure A3. Detection rates of LV dwarf galaxies in the extended source detection step within pre-process.

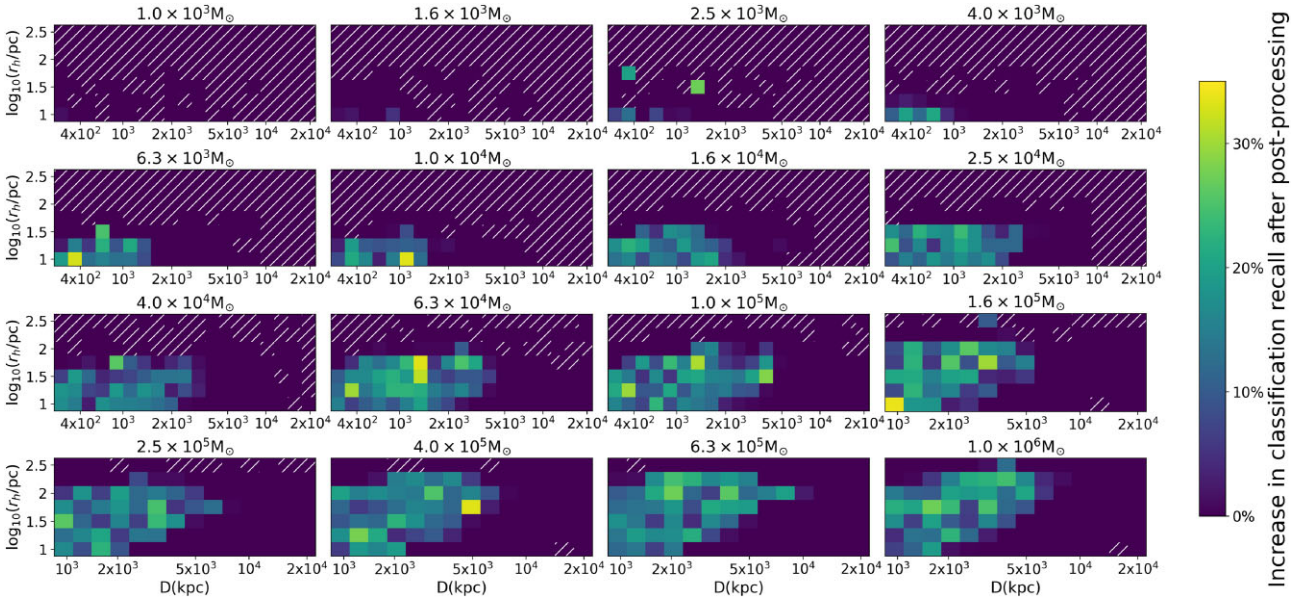

Figure A4. Enhancement of the classification recall through post-process.

This paper has been typeset from a TEX/LATEX file prepared by the author.