NATIONAL CENTRE FOR NUCLEAR RESEARCH

DOCTORAL THESIS

Identification and characterization of strong gravitational lenses and low surface brightness galaxies using deep learning

Author:
Hareesh THURUTHIPILLY

Supervisor:
prof. dr. hab. Marek Biesiada
Auxiliary supervisor:
dr. Junais

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

in the

Department of Fundamental Research, Astrophysics Division

June 17, 2024

Declaration of Authorship

I, Hareesh THURUTHIPILLY, declare that this thesis titled, "Identification and characterization of strong gravitational lenses and low surface brightness galaxies using deep learning" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at the National Centre for Nuclear Research.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at the National Centre for Nuclear Research or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

NATIONAL CENTRE FOR NUCLEAR RESEARCH

Abstract

Identification and characterization of strong gravitational lenses and low surface brightness galaxies using deep learning

Hareesh THURUTHIPILLY

Despite the success of the ΛCDM model over the past two decades, the model has encountered several unresolved tensions with observations. These include the nature of dark energy and dark matter, the missing baryon problem, and galaxy-scale tensions, collectively known as small-scale problems. To resolve these tensions, we need to test the predictions of the ΛCDM model on galaxy systems, which could challenge and refine our understanding of the Universe. Strong gravitational lenses (SGLs) and low surface brightness galaxies (LSBGs) are two such sources that we could use to resolve the tension of the ΛCDM model with observations. SGLs are systems formed when the gravitational field of a massive foreground galaxy distorts the image of a background galaxy or quasar to produce multiple images, arcs, or rings of the background source. SGLs can be used to constrain the cosmological models and estimate cosmological parameters. LSBGs are generally defined as galaxies that are fainter than the night sky, making them difficult to detect. Current predictions estimate that LSBGs occupy a significant fraction of the total galaxy population. Upcoming large-scale surveys, such as the Legacy Survey of Space and Time (LSST) and Euclid, are expected to identify about 10^5 SGLs and more than 10^5 LSBGs in the coming years, paving the way for the big data era in astronomy. However, the identification of these systems is challenging due to the rarity of SGL systems and the difficulty of distinguishing LSBGs from artefacts. In this thesis, inspired by the success of deep learning models in analysing everyday images, I introduce transformer models for identifying SGLs and LSBGs from large-scale surveys and successfully implement these models for this purpose.

To compare the performance of the transformers with CNNs for the identification of SGLs, I constructed and trained 21 transformer models and five CNNs to identify gravitational lenses from the simulated dataset of the Bologna lens challenge. I used four different metrics for evaluation: classification accuracy, the area under the receiver operating characteristic (AUROC) curve, and TPR0 and TPR10 scores (two metrics of evaluation for the Bologna challenge). I compared the performance of the transformer models and the best-performing models from the Bologna lens challenge and found that the transformer models performed better than all the models that participated in the challenge, including the CNNs that won the challenge. Transformer models can identify SGL candidates with a high level of confidence and will be able to filter out potential candidates from real data. I tested the transformer models on the SGL candidates found by the SGL searches in the Kilo Degree Survey (KiDS) and despite not being trained on the KiDS data, the transformer models were able to identify 65% of the SGL candidates.

In my next work, I study the use of transformer models in separating LSBGs from artefacts in the data from the Dark Energy Survey (DES) Data Release 1. I created eight different transformer models which achieved an accuracy of $\sim 94\%$ and used two ensembles of these eight models to identify LSBGs from DES. Using the transformer models, I then searched for new LSBGs from the DES that the previous searches may have missed and found 4 083 new LSBGs that had not been reported in the previous searches. Subsequently, the properties of the newly found LSBGs are investigated, along with an analysis

of the properties of the total LSBG sample in DES. With the addition of 4 083 new LSBGs I increased the sample size of known LSBGs in DES by $\sim$ 17% and increased the number density of LSBGs in DES to 5.5 deg^{-2}. The new LSBG sample consists of mainly blue and compact galaxies. I performed a clustering analysis of the LSBGs in DES using an angular two-point auto-correlation function and found that LSBGs cluster more strongly than their high-surface-brightness counterparts. This effect is mainly driven by the red LSBG. I associated 1310 LSBGs with galaxy clusters and identified 317 ultra-diffuse galaxy candidates among them. Analysing the cluster-centric properties, I found that these cluster LSBGs in DES are getting bluer and larger in size towards the edge of the clusters when compared with those in the centre.

For my next work, I studied the scope of transfer learning for the identification of LSBGs. I trained two ensembles of transformer models with data from the DES which achieved an accuracy $\sim$ 95%. Subsequently, these models are tested on the data of the Abell 194 cluster acquired from targeted observations with the Hyper Suprime-Cam (HSC), which is two orders of magnitude deeper than the DES data. I identified a sample of 171 LSBGs, of which 87 are completely new from the HSC data and further classified 28 LSBGs as ultra-diffuse galaxies (UDGs). The number of UDGs in the Abell 194 cluster aligns with observations in the literature that the number of UDGs scales proportionally with the mass of the cluster. Analyzing the cluster-centric properties of the sample, I found that the LSBGs and UDGs near the cluster centre are brighter and have higher Sérsic index values compared to those in the outer regions. Additionally, the LSBGs near the cluster centre tend to be redder than those in the outer parts of the Abell 194 cluster, with this trend being more pronounced in $FUV - NUV$ and $NUV - r$ colors compared to the $g - r$ color. Examining the color-magnitude ($NUV - r$ vs M_r) space of LSBGs it is found that the majority of the NUV emitting LSBGs are in the blue cloud with only two LSBGs being red in color which both are massive galaxies.

I have shown that the transformer models have the potential to be on par with CNNs as state-of-the-art algorithms in identifying SGLs and LSBGs. In addition, I have shown that transfer learning from a shallow survey to a deeper survey using transformer models could be successfully achieved. The methodology, I have developed as part of this thesis could prove valuable for identifying and analyzing astronomical data in upcoming surveys like LSST and Euclid.

NATIONAL CENTRE FOR NUCLEAR RESEARCH

Streszczenie

Identification and characterization of strong gravitational lenses and low surface brightness galaxies using deep learning

Hareesh THURUTHIPILLY

Pomimo sukcesu modelu ΛCDM w ciągu ostatnich dwóch dekad, model ten napotkał kilka nierozwiązanych napięć z obserwacjami. Obejmują one naturę ciemnej energii i ciemnej materii, problem brakujących barionów oraz niezgodności w skali galaktyk, zwane łącznie problemami małych skal. Aby rozwiązać te napięcia, musimy przetestować przewidywania modelu ΛCDM na układach galaktyk, co może stanowić wyzwanie i udoskonalić nasze rozumienie Wszechświata. Silne soczewki grawitacyjne (SGL) i galaktyki o niskiej jasności powierzchniowej (LSBG) to dwa takie źródła, które możemy wykorzystać do rozwiązania napięć w modelu ΛCDM za pomocą obserwacji. SGL to systemy powstające, gdy pole grawitacyjne masywnej galaktyki pierwszego planu zniekształca obraz galaktyki tła lub kwazara, tworząc wiele obrazów, łuków lub pierścieni źródła. LSBG są ogólnie definiowane jako galaktyki, które są bledsze niż nocne niebo, co czyni je trudnymi do wykrycia. Oczekuje się, że przyszłe wielkoskalowe przeglądy, takie jak Legacy Survey of Space and Time (LSST) i Euclid, w nadchodzących latach zidentyfikują około 10^5 SGL i ponad 10^5 LSBG, torując drogę do ery dużych zbiorów danych w astronomii. Jednak identyfikacja tych systemów stanowi wyzwanie ze względu na rzadkość systemów SGL i trudność w odróżnieniu LSBG od artefaktów. W niniejszej rozprawie, zainspirowany sukcesem modeli głębokiego uczenia się w analizie zwyczajnych obrazów, wprowadzam modele transformatorowe do identyfikacji SGL i LSBG z przeglądów na dużą skalę i z powodzeniem wdrażam te modele do tego celu.

Aby porównać wydajność transformatorów z CNN do identyfikacji SGL, skonstruowałem i wytrenowałem 21 modeli transformatorów i pięć CNN do identyfikacji soczewek grawitacyjnych z Bologna Lens Challenge. Do oceny wykorzystałem cztery różne wskaźniki: dokładność klasyfikacji, obszar pod krzywą charakterystyki operacyjnej odbiornika (AUROC) oraz wyniki TPR0 i TPR10 (dwa wskaźniki oceny dla wyzwania bolońskiego). Porównałem wydajność modeli transformatorowych oraz najlepszych modeli z wyzwania bolońskiego i stwierdziłem, że modele transformatorowe wypadły lepiej niż wszystkie modele, które wzięły udział w wyzwaniu, w tym również CNN, które wygrały to wyzwanie. Modele transformatorowe mogą identyfikować kandydatów na SGL z wysokim poziomem ufności i będą w stanie odfiltrować potencjalnych kandydatów z rzeczywistych danych. Przetestowałem modele transformatorowe na kandydatach na SGL znalezionych przez wyszukiwanie SGL w Kilo Degree Survey (KiDS) i pomimo tego, że nie zostały one przeszkolone na danych KiDS, modele transformatorowe były w stanie zidentyfikować 65% kandydatów na SGL.

W mojej kolejnej pracy badam wykorzystanie modeli transformatorowych do oddzielania LSBG od artefaktów w danych z Dark Energy Survey (DES) Data Release 1. Stworzyłem osiem różnych modeli transformatorowych, które osiągnęły dokładność $\sim$ 94% i wykorzystałem dwa zespoły tych ośmiu modeli do identyfikacji LSBG z DES. Korzystając z modeli transformatorowych, wyszukałem nowe LSBG z DES, które mogły zostać pominięte w poprzednich wyszukiwaniach i znalazłem 4083 nowe LSBG w gromadach galaktyk,

które nie zostały zgłoszone w poprzednich poszukiwaniach. Następnie zbadano właściwości nowo znalezionych LSBG w gromadach, wraz z analizą całej próbki LSBG w gromadach w DES. Dodając 4083 nowe LSBG w gromadach, zwiększyłem rozmiar próbki znanych LSBG w gromadach w DES o $\sim$ 17% i zwiększyłem gęstość liczbową LSBG w gromadach w DES do 5,5 $\deg^{-2}$. Nowa próbka LSBG składa się głównie z niebieskich i zwartych galaktyk. Przeprowadziłem analizę skupień galaktyk LSBG w DES przy użyciu funkcji autokorelacji kątowej dwupunktowej i odkryłem, że galaktyki LSBG grupują się silniej niż ich odpowiedniki o dużej jasności powierzchniowej. Efekt ten jest spowodowany głównie przez czerwone LSBG w gromadach. Powiązałem 1310 LSBG z gromadami galaktyk i zidentyfikowałem wśród nich 317 kandydatów na galaktyki ultra-rozproszone. Analizując właściwości gromad, odkryłem, że te LSBG w gromadach w DES stają się coraz bardziej niebieskie i większe w kierunku krawędzi gromad w porównaniu z LSBG w centrum.

W mojej następnej pracy badałem zakres uczenia transferowego do identyfikacji LSBG. Wyszkoliłem dwa zespoły modeli transformatorów z danymi z DES, które osiągnęły dokładność $\sim$ 95%. Następnie modele te zostały przetestowane na danych gromady Abell 194 uzyskanych z ukierunkowanych obserwacji za pomocą Hyper Suprime-Cam (HSC), które są o dwa rzędy wielkości głębsze niż dane DES. Zidentyfikowałem próbkę 171 galaktyk LSBG, z których 87 jest zupełnie nowych na podstawie danych HSC, a następnie sklasyfikowałem 28 galaktyk LSBG jako galaktyki ultra-rozproszone (UDG). Liczba UDG w gromadzie Abell 194 jest zgodna z obserwacjami w literaturze, zgodnie z którymi liczba UDG skaluje się proporcjonalnie do masy gromady. Analizując właściwości próbki w centrum gromady, stwierdziłem, że gromady LSBG i UDG w pobliżu centrum gromady są jaśniejsze i mają wyższe wartości indeksu Sérsic'a w porównaniu do LSBG w regionach zewnętrznych. Dodatkowo, LSBG w pobliżu centrum gromady mają tendencję do bycia bardziej czerwonymi niż te w zewnętrznych częściach gromady Abell 194, przy czym trend ten jest bardziej wyraźny w kolorach $FUV - NUV$ i $NUV - r$ w porównaniu do koloru $g - r$. Analizując przestrzeń barwno-magnitudową ($NUV - r$ vs M_r) LSBG stwierdzono, że większość LSBG emitujących NUV jest w kolorze niebieskim, a tylko dwie LSBG są w kolorze czerwonym, przy czym obie są masywnymi galaktykami.

Wykazałem, że modele transformatorowe mają potencjał, aby dorównać CNN jako najnowocześniejsze algorytmy w identyfikacji SGL i LSBG. Ponadto wykazałem, że uczenie transferowe z płytkiego przeglądu do głębszego przeglądu przy użyciu modeli transformatorowych może być z powodzeniem osiągnięte. Metodologia, którą opracowałem w ramach tej pracy, może okazać się cenna przy identyfikacji i analizie danych astronomicznych w nadchodzących przeglądach, takich jak LSST i Euclid.

Acknowledgements

Finishing my PhD was not an easy task for me. There were days when I felt super confident and days when I felt completely clueless.

I want to thank my supervisor Marek Biesiada who saw potential in me and accepted me to the NCBJ graduate school program as a PhD student. I am really grateful to my co-supervisor Junias for helping me whenever I got stuck during my PhD journey and patiently answering my queries. Without his help, this thesis would not have been a reality and thanks for being a good friend. I want to thank Agnieszka Pollo for the support and guidance she has given me during the last 4 years. Her insights have helped me understand what it means to be a good scientist, and thanks for setting up an example for me to look forward to. I also want to thank Jin Koda, Masafumi Yagi, Yutaka Komiyama, and Yamanoi Hitomi for giving me access to the data of their dedicated HSC observations, which helped me write the last chapter. In addition, I would like to thank my collaborators Unnikrishnan and Samuel Boissier for the discussions and insights they have given me.

I want to thank my mother Bindu and father Haridas for their unwavering support and belief in me. Without the support from both of them, I would have never been able to start this journey. I want to thank my sister Haritha for her support and my cousin Suni for encouraging me to dream higher and bigger and believing that I could reach those heights. I also want to thank my class teacher P.V. Chacko from 8th standard who made me believe in myself and made me confident to start the academic journey.

The past four years were made enjoyable by the constant support of my friends. I am particularly grateful to Sajal, Jefry and Aswin for standing by me during difficult times. I also want to thank Linson, Yadhu, Ansa, Adil, Athira, Akash, Jaimon, and Malu for their support. My friends and colleagues at NCBJ, including Yashwanth, Prasad, Margherita, Sreekanth, Shuaibo, and everyone from the NCBJ grad school, made the stressful times bearable, and I thank you all for your support.

Contents

List of Figures

List of Tables

All civilisations have an origin myth. We are the first to get it right - David Tong.

1

Introduction

From time immemorial, humans have wondered about our status on the vast scale of the cosmos and pondered questions such as what is our origin and what is our ultimate fate. Not surprisingly, these questions have been the subject of religious and philosophical speculation, and humans have always looked at astronomical bodies with curiosity to answer them. Observational astronomy and cosmology are the results of our scientific endeavour to answer these questions. Owing to technological advancements such as charge-coupled device (CCD) cameras and pioneering breakthroughs in our theoretical understanding of the cosmos, namely via general relativity, we have been able to unravel the cosmos on an exponential scale in the last 20th century. However, we still have loose ends, and this thesis is focused on ways to navigate in the era of big data to tighten some of the loose ends.

1.1 Introduction to Cosmology

Compared to other areas of physics, cosmology could be considered a very new discipline. Shortly after Einstein (1915) established the theoretical framework of general relativity, he explored the solutions of the field equations to the universe (Einstein 1917). To his surprise, all of the solutions to the universe described either an expanding or a contracting universe, contrary to the popular belief that the universe was static. To make the universe static, Einstein introduced a cosmological constant to his field equations, which would counteract the expansion and make the universe static.

In the same period, Alexander Friedmann, Georges Lemaitre, Howard P. Robertson, and Arthur Geoffrey Walker independently explored the solutions to Einstein's field equations for our universe (Friedmann 1924; Lemaître 1927; Robertson 1935). The solutions they have found are known as the Friedmann-Lemaître-Robertson-Walker metric, or FLRW metric, which describes a homogenous and isotropic universe undergoing expansion. The concept that the cosmos is homogenous and isotropic may seem counter-intuitive, as it is evident from our everyday experiences that the world we inhabit is not homogenous or isotropic. Nevertheless, the solutions to the field equations were derived with the premise that the cosmos will exhibit homogeneity and isotropy on extremely large scales, a concept referred to as the cosmological principle. The early solutions of the universe assumed that the major components of the universe were baryonic matter and radiation.

1.1.1 Evidence for Dark matter

Zwicky was the first to introduce the idea of dark matter in 1933 when he was analysing the rotational velocities of galaxies in the Coma Cluster. His analysis found that the total

matter needed to hold the cluster together, as found with the virial theorem, was 400 times larger than the mass of the luminous matter found in the cluster (Zwicky 1933). He called it *dunkle Materie*, which is German for dark matter. However, the idea of the existence of dark matter was not completely accepted until the 1970s, which is 40 years after the discovery of Zwicky. Ostriker et al. (1974) and Einasto et al. (1974) found that the very massive halos are required to support the motion of the satellite galaxies around the Milky Way as well as for the nearby galaxies. Similarly, the lack of falloff in the rotational velocities of the spiral galaxies at larger radii also indicated the presence of non-visible matter extended beyond visible boundaries of the galaxy (Roberts & Rots 1973; Rubin et al. 1978, 1980). These results pave the pathway to the idea that the galaxies are embedded in large dark matter halos and an illustration of the Milky Way galaxy embedded in the dark matter halo is shown in Fig. 1.1.

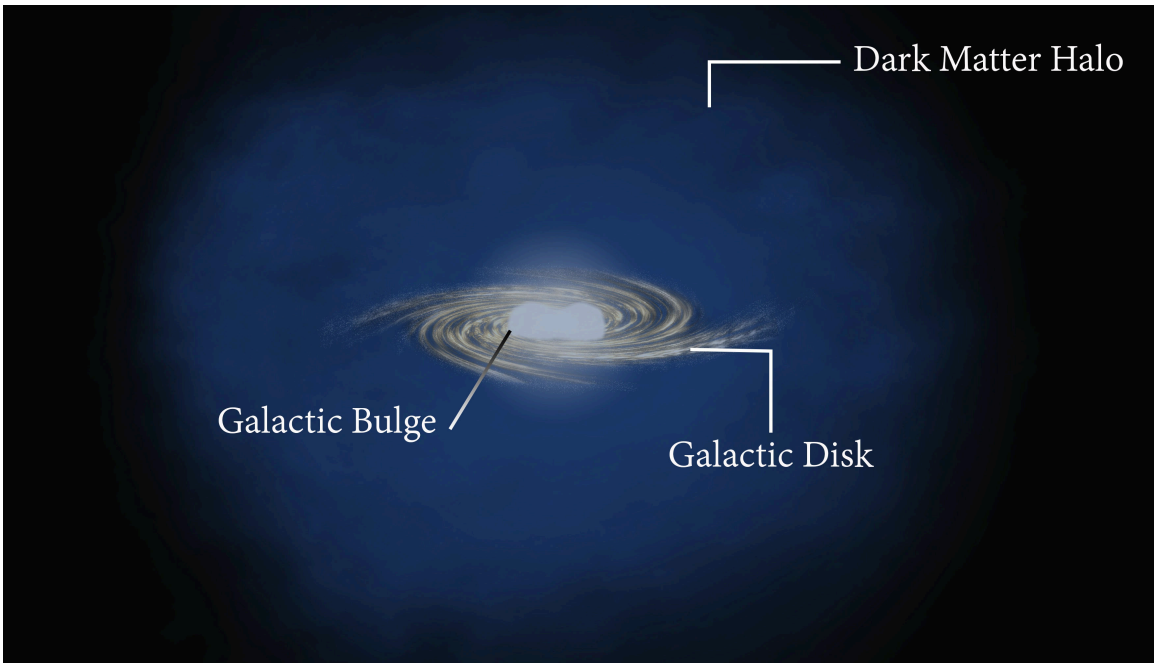


FIGURE 1.1: Illustration of the Milky Way galaxy embedded in the dark matter halo. Image credit: L. Jaramillo & O. Macias(Virginia Tech)

In the same period, numerous candidates were proposed to account for this missing matter, such as brown dwarfs, white dwarfs, black holes and neutrinos. However, the exact nature of dark matter is still debated. Neutrino-based dark matter candidates or particles with a mass less than 30eV are generally called hot dark matter (HDM) because of their velocity being comparable to the speed of light (Bond et al. 1980; Sato & Takahara 1980). However, HDM models could not correctly explain the formation of large-scale structures in the universe (Tremaine & Gunn 1979; White et al. 1984). With the introduction of supersymmetry in the 1980s and the development of various particle physics-based models, there were numerous candidates for dark matter, such as photinos, neutralinos, and gravitinos which were non-relativistic particles. As a result, the class of candidates with non-relativistic velocity is generally known as cold dark matter (CDM) (Peebles 1982). It should be noted that the distinction between HDM and CDM is based on their relativistic nature. Simulations show that the large-scale structure forms hierarchically in a universe with CDM, which is in excellent agreement with observations (Davis et al. 1985).

1.1.2 Accelerated Expansion of the Universe

In 1929, Edwin Hubble showed that the recession velocities of galaxies are linearly proportional to the distance ($v \propto r$) to the galaxy (Hubble 1929). This relation is nowadays known as Hubble–Lemaître law and can be mathematically represented as $v = H(t)r$. It is important to note that the proportionality constant, known as Hubble's constant, varies with time but remains constant in space. This indicated that all the galaxies are moving away from each other and supported the argument that we are living in an expanding universe. The news of the expanding universe made Einstein announce that the introduction of the cosmological constant was the greatest blunder in his life. However, the cat was out of the bag, and in the present age, the cosmological constant solves a different problem and creates another.

If the universe is expanding, it also logically follows that in the past, the universe was in a state of high density and high temperature. This led George Gamow to predict that the elements observed in the universe might be synthesised through thermonuclear reactions during the early stages of the universe, referred to as primordial nucleosynthesis (Gamow 1946, 1948; Alpher et al. 1948). Another prediction stemming from this concept was the existence of residual heat radiation from the early universe that would persist in the present universe (Alpher et al. 1948). Fred Hoyle mockingly referred to the idea that the universe originated from a dense and hot starting state as the Hot Big Bang. However, the accidental discovery of cosmic microwave background radiation (CMBR) by Penzias & Wilson (1965) validated the Bing Bang model of cosmology.

By the end of the 20th century, the idea that the universe was expanding was widely accepted. However, it was anticipated that the recession velocity, at which galaxies move apart, would be slowing down due to the gravitational forces of attraction. Since the galaxies are moving away from us, the radiation from these galaxies will be redshifted, and measuring the redshift in the radiation can be used as a proxy for the recession velocities of the galaxies. Nevertheless, accurately determining the distance to the galaxies poses a significantly more challenging issue. Since the intrinsic brightness of supernova-Ia (SN-Ia) is expected to be constant, measuring the apparent brightness of SN-Ia could be used to determine the distance to the explosion. Based on this idea, Riess et al. (1998) and Perlmutter et al. (1999) independently analysed the light curves from the SN-Ia and found that the universe is not decelerating and, on the contrary, is undergoing an accelerated expansion. The accelerated expansion of the universe points to a strictly positive cosmological constant corresponding to a mysterious energy component known as dark energy. The exact nature of dark energy is still debated and remains one of the unsolved problems in physics.

1.1.3 Λ CDM Cosmology

A universe described by a positive cosmological constant and with dark matter as CDM is known as ΛCDM universe. The detection of cosmic microwave background radiation turned a new leaf in astronomy as it gave a way to probe the early universe. Independent measurements of the acoustic peaks and their amplitude and wavelength in the CMB power spectrum showed that they were in excellent agreement with the predictions of ΛCDM cosmology (de Bernardis et al. 2000; Hanany et al. 2000; Halverson et al. 2002). Similarly, precise measurements of the cosmological variables estimated from the CMB data with the Wilkinson Microwave Anisotropy Probe (WMAP) were in good agreement with independent measurements of these quantities (Hinshaw et al. 2013a). For instance, the baryon density estimated with cosmological nucleosynthesis and the dark energy density content estimated from the SN-Ia were consistent with the measurements of the same

with CMB (Spergel et al. 2007). These results made the ΛCDM model the current standard model of cosmology.

One of the remarkable results of the ΛCDM model is the prediction of structure formation that is consistent with the large-scale structures observed in the current universe. In the very early stages of the universe with CDM, both dark matter and baryonic matter were coupled with radiation. As cosmic time progresses, the universe begins to cool down, and the dark matter decouples from the radiation faster than the baryons so that their density perturbation can grow. Following the decoupling of baryons from radiation, they rapidly fall into these dark matter potential wells. Thereby starting the formation of structures at a sufficiently rapid rate to align with the presently observed large-scale structure. Davis et al. (1985) used simulations to show that the present large-scale structures could be well explained by CDM models, and even since the 1990s, CDM models have been considered the best-fit model for describing dark matter.

Nevertheless, it should be noted that the ΛCDM model is only a best-fit model of the universe, and there could be a better model for the universe that is waiting to be discovered or tested. The ΛCDM model was successful in explaining the components of the universe and the distribution of large-scale structures. However, in the scientific realm, a few contradicting experiments or observations are needed to doubt the legitimacy of the theory. In the case of the ΛCDM model, we also have some problems that need to be addressed and resolved. For a detailed discussion and review, please refer to Perivolaropoulos & Skara (2022) or Di Valentino et al. (2021). One of the probes to test the consistency of a cosmological model is to observe how the galaxies are distributed in the cosmos and study their properties to check with the predictions from the ΛCDM model.

1.2 World of Galaxies

Having developed powerful telescopes by the end of the 18^{th} century, astronomers have understood that not all the intrinsically bright objects in the sky are stars or comets. These objects were fuzzy rather than point-like objects, as stars would be, and were collectively called nebulae. From 1771 to 1784, Charles Messier compiled a catalogue of these nebulae with more than 100 sources so as not to get confused with comets. John Herschel conducted the first organized search for nebulae, and in 1864, he published a list of 5079 objects that resembled nebulae (Mo et al. 2010).

Even after many years of its discovery, the debate about whether the nebulae were part of the Milky Way or not continued. One hypothesis was that all nebulae are celestial bodies situated within our Milky Way, while the other suggested that certain nebulae might be extragalactic entities, functioning as distinct "island universes" similar to the Milky Way. In 1925, Edwin Hubble solved this dispute once and for all by measuring the distance to the cephid variable stars and concluded that some nebulae are extra-galactic and have the same size and luminosity as our host Milky Way galaxy (Hubble 1925). This discovery began the journey of observational extra-galactic astronomy, and currently, more than 10^7 galaxies similar to our Milky Way galaxy have been catalogued compared to the 100 nebulae sampled by Charles Messier.

As per our current understanding, the fundamental building blocks of the large-scale structures that we see in our universe are galaxies, which are gravitationally bound systems of stars, stellar remnants, dust, interstellar gas, and dark matter. Analysing the formation and evolution of galaxies is intended to test our understanding of physics across vast scales in terms of size and time. Studying galaxy evolution and formation is distinct from other areas of experimental physics due to the significant disparity in timescales. Even the shortest timescales in galaxy evolution models are considerably longer than the

average duration of human civilisations. However, due to the finite speed of light, observing galaxies that are farther away is essentially equal to observing galaxies at a younger phase in the universe. Hence, in a statistical sense, one can study galaxy evolution and formation by comparing the properties of the galaxies at different epochs of the universe and also studying the relations between the galaxies in each epoch.

1.2.1 Observational tools in astronomy

Galaxies are a diverse class of astronomical objects; consequently, we need numerous parameters to characterise or represent a galaxy. Not surprisingly, all the information about a galaxy is derived from the radiation it emits at different wavelengths and from measuring the flux of the radiation. Due to historical reasons, the flux from an astronomical object as observed by an observer is represented by the apparent magnitude, which is given by:

$$m_x = -2.5 log_{10}(F_x / F_{x,0}) \tag{1.1}$$

where F_x is the measured flux in the photometric filter x and $F_{x,0}$ is the reference flux (zero-point) for that photometric filter (Mo et al. 2010). By definition, the apparent magnitude is reverse logarithmic; hence, the brighter an object, the lower its magnitude. In addition, the apparent magnitude also depends on the distance between the astronomical object and the observer. To represent the intrinsic brightness of an astronomical object, one needs to measure its absolute brightness (M_X), which is the apparent magnitude it would possess if it were observed from a distance of 10 parsecs. Here, a parsec ($\sim 3 \times 10^{16} m$) is defined as the distance between the Sun and an astronomical object that has a parallax angle of one arcsecond and is often used as a unit for distance in astronomy.

The absolute magnitude and the apparent magnitude are related by the relation:

$$M_x = m_x - 5(log_{10} d_L - 1) \tag{1.2}$$

$$d_L = \sqrt{L / 4\pi F_x}. \tag{1.3}$$

Where d_L is the luminosity distance to the astronomical object and L is the intrinsic luminosity of the object (Mo et al. 2010). By definition, one needs to know the luminosity distance of an astronomical object to measure its absolute magnitude. Similarly, if one knows the M_x and m_x, which are often combined to define the distance modulus ($M_x - m_x$) of an astronomical object, one could determine the luminosity distance to that astronomical object. These relations are fundamental in astronomy to investigate and study galaxy evolution and cosmology.

The galaxies also come in different colours. Generally, in astronomy, the color of a galaxy is determined by the ratio of its luminosity in two different photometric passbands. In practice, the color of a galaxy between two bands can be found by subtracting the magnitude in the red filter band from those in the blue filter band (blue-red), owing to the logarithmic relationship between magnitudes. In practice, a population of galaxies can generally be attributed to a bimodality in color and is typically classified into either the red or blue population. A galaxy is classified as red if its luminosity in the redder filter band is much greater than its luminosity in the bluer filter band. Blue galaxies typically denote highly active, star-forming spiral or irregular systems, while red galaxies predominantly represent spheroidal or elliptical formations (Strateva et al. 2001).

In addition to the magnitudes, astronomers often also use the surface brightness of a galaxy. The surface brightness of an astronomical object is defined as the apparent brightness per angular area and resembles how luminous matter is distributed in the galaxy. Knowing the apparent magnitude (m), and the surface area (A) occupied in the sky by the

object, one can calculate the mean surface brightness as (Tanoglidis et al. 2021b):

$$\bar{\mu} = m + 2.5log_{10}(A). \tag{1.4}$$

For observing an astronomical object, the surface brightness might be more informative than the magnitude. Because it tells how much the light from the galaxy is distributed in the galaxy, and if it has a higher concentration, then it would be easy to spot in the sky.

Generally, the variation of the surface brightness of a galaxy as a function of the distance from its centre can be described by the mathematical function known as Sérsic profile (Sérsic 1963). The Sérsic profile in terms of surface brightness is given as (Caon et al. 1993):

$$\mu(R) = \mu_e + \frac{2.5}{ln10} b_n \left[\left(\frac{R}{R_e} \right)^{1/n} - 1 \right] \tag{1.5}$$

where R_e is the radius at which half of the total light from the galaxy is contained, which is also known as the half-light radius, μ_e is the surface brightness at the half-light radius and n is known as the Sérsic index. The constant b_n is chosen such that it satisfies the condition $\Gamma(n) = \gamma(2n, b_n)$ where Γ and γ are the complete and incomplete gamma functions. The variation of surface brightness as a function of distance from the centre with fixed R_e in arbitrary units is shown in Fig. 1.2. It should be noted that when the Sérsic index value n increases, the centre core becomes steeper and the surface brightness outside of R_e becomes more extended. Whereas a smaller value of n results in a flatter surface brightness distribution in the core and has a sharper cut-off in the surface brightness outside R_e. For a detailed discussion of the surface brightness profiles of galaxies, please refer to Graham & Driver (2005).

1.2.2 Galaxy Classifications

The Sérsic index is also found to be correlated to the galaxy morphology. Historically, the first attempt to classify galaxies based on their morphology was made by Hubble (1926) using the images from the photographic plates. Hubble (1926) divided the galaxies into four different morphologies-elliptical (E), spiral (S), lenticular (S0 and SB0), and irregular (Irr). The elliptical galaxies were again subdivided based on their ellipticities, and the spiral galaxies were divided based on the presence of bars or not (S and SB). This classification scheme is known as the Hubble tuning fork because of the shape in which it is usually represented. An illustration of the Hubble tuning fork is shown in Fig. 1.3.

Generally, the discs of spiral galaxies are observed to host many young stars and regions of active star formation, while elliptical galaxies have predominantly old stellar populations. Early-day astronomers hypothesised that the galaxy classification resembling a tuning fork is a consequence of galaxy evolution and labelled elliptical and lenticular galaxies as early-type galaxies, and spiral and irregular galaxies were labelled as late-type galaxies. This arrangement, depicted in the schematic known as the Hubble tuning fork or sequence, was later discovered to be inaccurate, and as per our current understanding, the early universe consisted of more spiral and irregular galaxies than elliptical galaxies (Kuhn et al. 2023). However, the classification persists, labelling irregular and spiral galaxies as late-type, while lenticular and elliptical galaxies are termed early-type galaxies.

Elliptical galaxies generally have smooth surface brightness distributions and have little amounts of cold gas or dust. Usually, the elliptical galaxy population tends to be redder in color indicating that the stellar population in the elliptical galaxies is older. Similarly, the surface brightness profiles of bright elliptical galaxies are found to be best described by the Sérsic index, $n = 4$ which is also known as the de Vaucouleurs profile (de Vaucouleurs 1948). However, this could not be generalised as the n is strongly correlated with the size

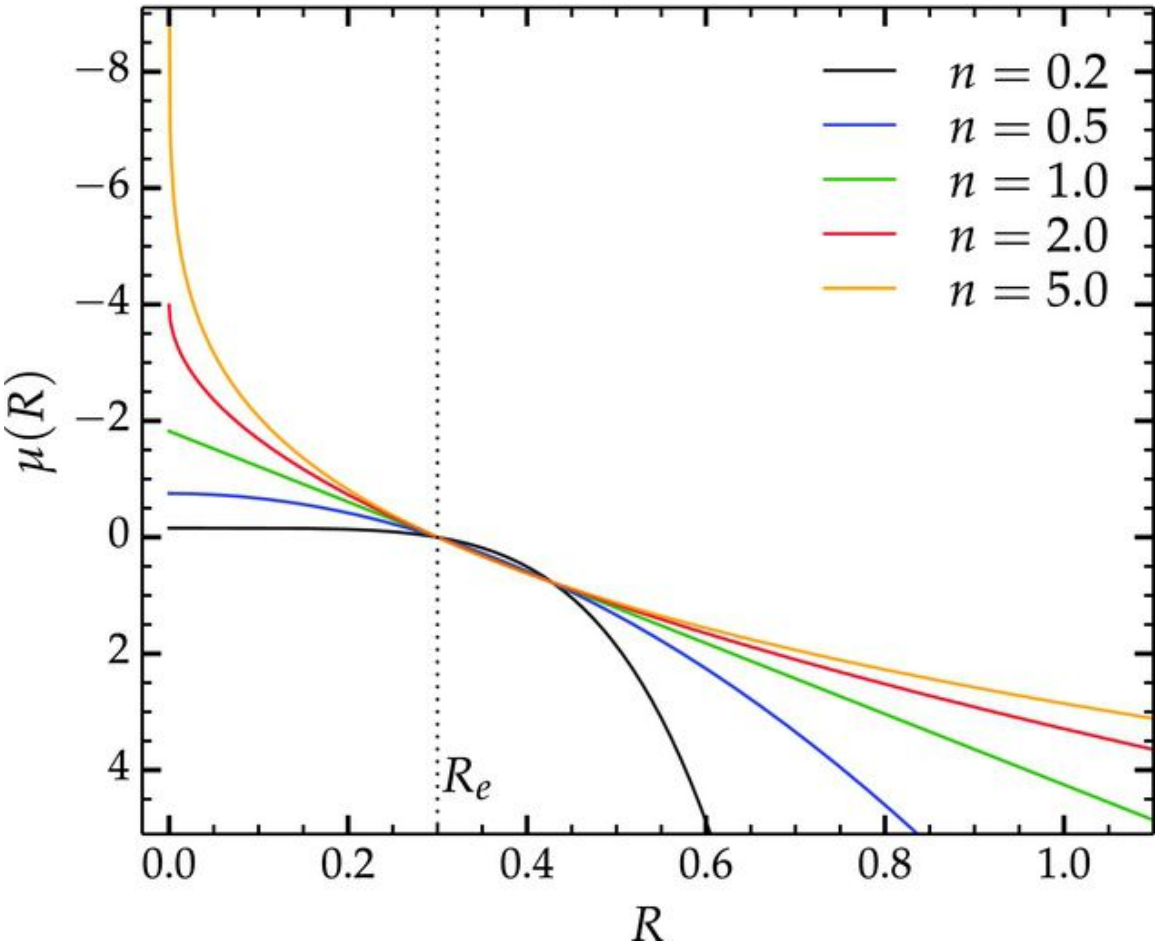


FIGURE 1.2: The Sérsic profile with different Sérsic index and same half-light radius (R_e). R_e is represented by a vertical dotted line. Image credits: Ciambur (2016)

and luminosity of the galaxy (Graham & Guzmán 2003). The faint elliptical galaxies are found to have $n \sim 0.5$ whereas the best fit for bright elliptical galaxies is found to have n as high as 10 (please see figure 2.13 from Mo et al. 2010). In general, the statistical dispersion of velocities attributed to the stars within an elliptical galaxy, known as the central velocity dispersion (σ_0), is found to be correlated with the mass of the galaxy as well as its size. This correlation is best described by the relation $logR_e = alog\mu_e + blog\sigma_0$ (Djorgovski & Davis 1987), where a and b are proportionality constants and μ_e is the mean surface brightness within the radius R_e.

Spiral galaxies have far more complexity in nature compared to elliptical galaxies. Usually, they have a thin, rotationally supported disc with spiral arms, a bar, and a central bulge component. Depending on the presence of a bar or not, they have been further subdivided into two different morphologies (S and SB) which in turn are subdivided based on distinct arm structures (see Fig. 1.3). Young stars, HII regions, molecular gas, and dust absorption are the main characteristics of the spiral structure and are most easily observed in face-on systems. The disks in the spiral galaxies are generally described by a surface brightness profile with an Sérsic index, $n \sim 1$ whereas the bulge component might be best described with a higher Sérsic index. Similar to elliptical galaxies, spiral galaxies also follow a scaling relation known as the Tully-Fischer relation, which states that the luminosity of a spiral galaxy is proportional to the rotation velocity (Tully & Fisher 1977).

The lenticular galaxies have similarities to elliptical and spiral galaxies and share the

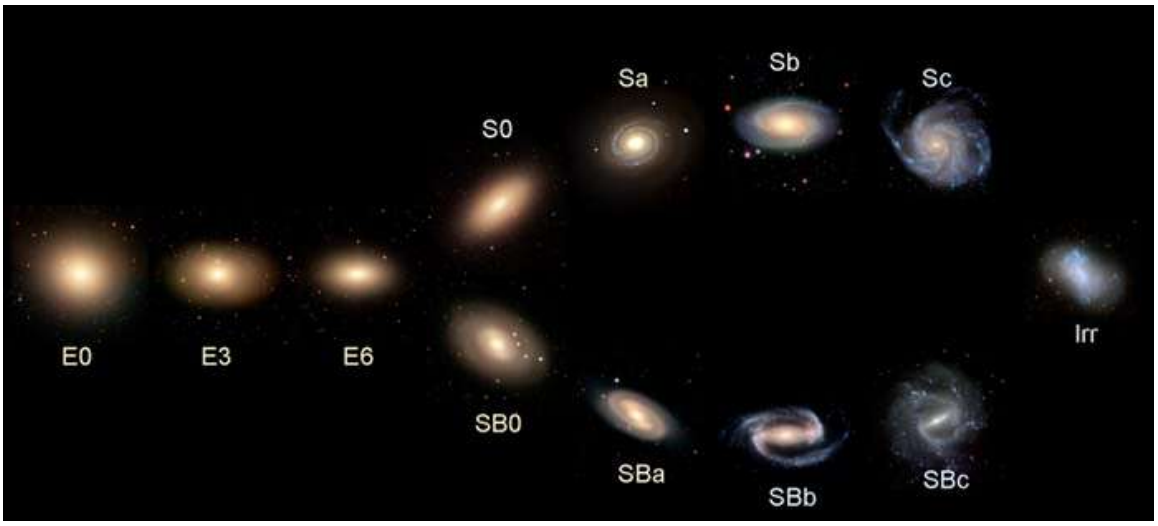


FIGURE 1.3: The Hubble tuning fork diagram. Image credits: Zooniverse.

characteristics of both of them. Similar to ellipticals, lenticulars lack spiral limbs and HII regions, resulting in a smooth light distribution and generally a redder population of stars. However, similar to spiral galaxies, they have a slender disc and a bulge in the centre, making them a distinct class compared to spiral or elliptical galaxies. The last class of galaxies, according to Hubble's classification, are characterised by the absence of a dominant bulge, a disc that is rotationally symmetric, and any discernible symmetry. In short, there is no regular pattern or structure for these galaxies, and hence they were named irregular galaxies. Hubble omitted this particular class from his initial sequence due to his uncertainty regarding whether it constituted an extension of the preceding classes. Presently, irregular galaxies are considered extensions of spiral galaxies. The classification scheme created by Hubble was later revised with the identification of more and more distinct morphologies, such as the peculiar galaxies. However, the general structure of the Hubble tuning fork is still used with revised morphologies to show the different galaxy morphologies.

It should also be noted that the galaxies could be classified based not only on their morphologies but also on their other properties, such as luminosity, surface brightness, gas content, the current level of star formation, and the presence of active nuclei in the centre of the galaxy.

1.2.3 Galaxy Surveys

Owing to the advancements in technology, we currently have the ability to observe the cosmos in various wavelength in the electromagnetic spectrum. Generally, astronomical surveys can be classified into two categories: spectroscopic and photometric surveys. In a spectroscopic survey, a high-resolution spectrograph is used to capture the spectrum of an astronomical object, which could be used to estimate its characteristics and redshift. If an object is moving away from us, then the wavelength of the light emitted by the object will shift to a higher wavelength, which is represented by the quantity known as redshift (z). The mathematical representation of redshift is

$$z = \frac{\lambda_{observed} - \lambda_{emitted}}{\lambda_{emitted}} = \frac{\lambda_{observed}}{\lambda_{emitted}} - 1 \tag{1.6}$$

where $\lambda_{observed}$ is the wavelength observed and $\Lambda_{emitted}$ is the wavelength of the emitted radiation. The redshift surveys can map the 3D distribution of the universe in great detail. Generally, the time taken to obtain the spectrum depends on the apparent brightness of the object, and the redshift surveys generally map the bright galaxies in the universe. Hence, their effectiveness is constrained by the sensitivity of spectrographs and the time and financial limitations required to obtain high-quality spectra from a large number of sources.

Photometric surveys, on the other hand, use wide-field cameras with different photometric filters to image large regions of the sky in detail. Hence, the photometric survey can detect very faint sources over a large scale, depending on the sensitivity of the camera. Photometric surveys image the sky in different wavelengths, encoding most of the information about the galaxy in these images which could be used to estimate the flux radiated by the sources in each wavelength. Multi-wavelength observations of a galaxy are important because the observations at each specific wavelength trace back to different physical processes. For instance, observing the sky at radio frequency helps us understand the presence of gas cool enough for molecules such as CO_2 and neutral hydrogen atoms to exist. Similarly, the main sources of emission in the infrared frequency range are low-mass stars as well as interstellar dust. The radiation in the optical regime comes from stars with masses similar to those of our sun, while highly massive stars (in comparison to the mass of the sun) emit ultraviolet radiation. High-energy radiation, such as X-rays, is emitted from the hot gas surrounding supermassive black holes as well as from the intracluster medium, which is a characteristic of galaxy clusters. The image of the Andromeda galaxy observed at various wavelengths is shown in Fig. 1.4, and here you can see that the structures traced in the image are different. Hence, multiwavelength observation is crucial to understanding the complete picture of a galaxy.

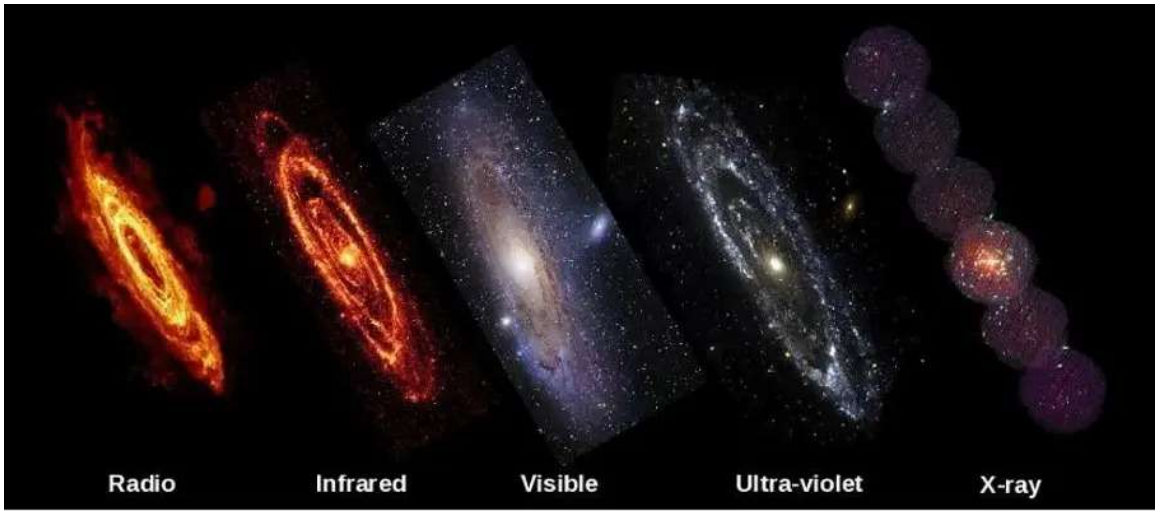


FIGURE 1.4: Image of the Andromeda galaxy as observed from radio to the X-ray. Credits: ESA / NASA.

As previously mentioned, the colors of galaxies provide valuable information about their stellar populations. Multi-wavelength observations allow us to estimate the colors of galaxies across different bands, each of which traces different physical phenomena. For instance, large-scale surveys typically cover the sky in the g, r, and i bands within the optical regime, with the $g - i$ color indicating the presence of an older stellar population. Additionally, an excess of UV radiation compared to optical or infrared radiation suggests recent star formation in the galaxy, as highly massive stars, which emit strongly in the UV, burn out more quickly.

In the early days of astronomical observations, because of the limited technology and resources, astronomers were only able to observe very small parts of the sky. As a consequence, sometimes the conclusions drawn from surveying such a small region of the sky were biased. Due to the limited resources, observing larger areas of the sky results in incompleteness in imaging the fainter structures in the sky. Currently, with improved technology, we have been able to develop powerful telescopes such as the Hubble Space Telescope (HST) and the James Webb Space Telescope (JWST). These modern telescopes can capture deep images of the local universe as well as observe galaxies from the early universe, providing unprecedented insights into cosmic history. However, due to technical limitations, these telescopes can only be used for targeted observation of small regions of the sky and can not be used for imaging a large region of the sky. Hence, generally, we need a trade-off between the completeness of the objects observed and the area of the sky observed to reach optimum results. To draw statistically confident conclusions about the properties of galaxies, we need larger samples of galaxies across different wavelengths.

Keeping this goal in mind, we have started imaging large parts of the sky at different wavelengths. For instance, the 2MASS has surveyed the whole sky in infrared wavelength (Skrutskie et al. 2006). Similarly, another notable survey of the sky is the Sloan Digital Sky Survey, which imaged $\sim$ 35% of the sky, observing $\sim 10^9$ objects and acquiring the spectra over 4×10^6 objects (Abazajian et al. 2009; Eisenstein et al. 2011; Ahumada et al. 2020). All these surveys at longer wavelengths were done with ground-based telescopes; however, to probe the universe at shorter wavelengths, such as ultraviolet or X-ray, one has to use space-based telescopes due to the atmospheric cut-off limits at these wavelengths. For instance, the Galaxy Evolution Explorer (GALEX), also known as Explorer 83 was a space telescope designed to study the universe in ultraviolet wavelengths (Martin et al. 2005; Bianchi et al. 2017). Its primary objective was to quantify the timeline of star formation across cosmic history. Similarly, with the ROSAT Survey, the entire sky was mapped in the X-ray regime, which helped us understand more about galaxy clusters and supermassive black holes (Voges et al. 1999; Boller et al. 2016). These large-scale observations help us understand how galaxies are distributed in space and about the environment of these galaxies.

One of the shortcomings of the earlier large-scale surveys was that they were not deep enough to detect faint sources and lacked completion in the faint end of the galaxy populations. For instance, the completeness of galaxies in the SDSS drops quickly for surface brightness greater than $\sim$ 24.5 mag arcsec^{-2} (Kniazev et al. 2004; Driver et al. 2005). However, this scenario has changed with the current optical surveys, such as the Kilo Degree Survey (KiDS), Dark Energy Survey (DES), and Hyper Suprime-Cam Subaru Strategic Programme. These modern surveys have pushed the boundaries, reaching depths of approximately 28 mag arcsec^{-2}. Nevertheless, this progress marks just the beginning. The forthcoming large-scale surveys, including the Legacy Survey of Space and Time (LSST) and Euclid, are designed to go even deeper. These surveys are anticipated to uncover structures with surface brightness around 30 mag arcsec^{-2} in almost half of the sky and will revolutionise the era of observational astronomy.

LSST and Euclid

The LSST is an upcoming ground-based large-scale imaging survey planned to cover 18 000 deg^2 of the sky in g, r, i, z, and y bands, with a surface brightness detection limit of around 30 mag arcsec^{-2} (Ivezić et al. 2019). Similarly, Euclid is a space-based large-scale survey that will observe 15 000 deg^2 of the sky in visible and near-infrared (NIR) wavelengths (Euclid Collaboration et al. 2022a). LSST and Euclid represent the next phase in large-scale surveys, surpassing their predecessors like DES and KiDS in both detection

limits and sky coverage. The area of the sky that will be observed by LSST and Euclid is shown in FIg. 1.5. In addition, it should be noted that both LSST and Euclid will cover the already observed regions of the sky by surveys such as DES and KiDS to extend our understanding.

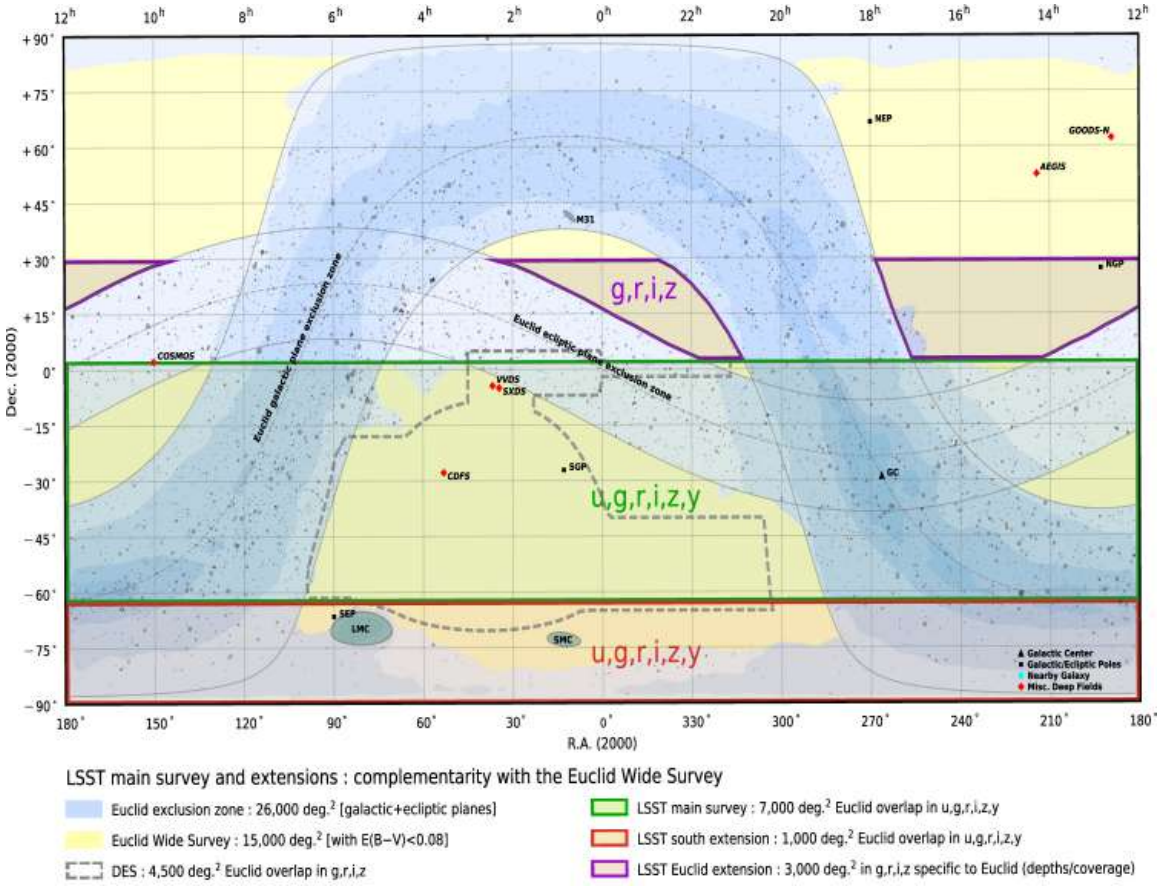


FIGURE 1.5: Footprint of DES, LSST, and Euclid. The area of the sky overlapping between Euclid and LSST is also marked. Credits: Rhodes et al. (2017).

The main goals of LSST and Euclid are to resolve some of the challenging questions faced by the ΛCDM model, such as the nature of dark energy and dark matter, by mapping the structure of the universe and studying the formation and evolution of galaxies.

1.3 Challenges of ΛCDM universe

In the previous two sections, I have described the ΛCDM cosmology and the properties of the galaxies that we observe in the visible universe. If the universe we live in is truly a ΛCDM universe, then all the observations that we see in the universe should be explained by the ΛCDM model. However, despite its undeniable success, the ΛCDM model has encountered some potential challenges in recent years. With the increasing availability of data from large-scale surveys and improved computing power enabling highly precise N-body simulations, tensions have begun to emerge between the observational data and the prediction of the model. For a detailed discussion on the challenges faced by the ΛCDM model, please refer to Perivolaropoulos & Skara (2022).

Nature of the dark sector and the lack of direct detection

Current estimates of the cosmological parameters predict that $\sim$ 70% of the universe is composed of dark energy and $\sim$ 25% is filled with dark matter. Despite being the major components of the universe, we do not have an exact idea about the nature of dark energy and dark matter, and we have not been able to detect them directly in any experiments (Merritt 2017). Dark energy is expected to be distributed uniformly over the universe and will not interact with matter through any fundamental interactions other than gravity. Attempts have been made to explain dark energy as the ground state energy in quantum field theories but have resulted in a discrepancy of the order of 120 magnitudes with the observed energy density. Even after an extensive search for particles that could be considered CDM, they have not been detected yet, and as for the case of dark energy, their predicted energy density is very low, making it almost impossible to detect them in laboratories.

Hubble tension

Another prominent tension that needs to be solved in the ΛCDM model is the Hubble tension. Independent measurements of the Hubble constant (H_0) with CMBR by Planck Collaboration et al. (2020) and using SN-Ia by Brout et al. (2022) have resulted in values that do not agree with each other in 3σ uncertainty. The measurement of H_0 by combining the data on baryon acoustic oscillations (BAO), clustering, and big bang nucleosynthesis (BBN) has resulted in a value that is in agreement with the measurements from CMBR (Abbott et al. 2018b). On the contrary, independent measurements of the Hubble constant with strong gravitational lenses (SGLs) also result in a Hubble constant that agrees with SN-Ia measurements but is in tension with the results from the CMBR (Wong et al. 2020). In short, the Hubble tension can be summarised as the disagreement in the measurement of the Hubble constant between early-universe probes (such as CMBR, BAO, BBN) and late-universe probes (such as SN-Ia, SGLs). A comparison of the measurements of the Hubble constant with different probes and the resulting Hubble tension is shown in Fig. 1.6. One of the possible solutions to the Hubble tension is to revise the cosmological model itself.

In cosmology, the relation between the pressure (p) and the density (ρ) of a component is known as the equation of state and is generally represented as

$$p = w\rho \tag{1.7}$$

where w will depend on the component. For matter and radiation, the value of w is 0 and $1/3$, respectively. If a cosmological constant describes dark energy, then the value of w is -1. However, there are arguments that the dark energy density might vary with time, so w will be a function of time (or redshift). Collectively, these models are known as quintessence models (Tsujikawa 2013). Independently constraining the equation of the state of dark energy could help us alleviate the Hubble tension. For a detailed discussion on the Hubble tension and its possible solutions, please refer to Di Valentino et al. (2021).

Missing Baryons

The measurements of anisotropies in the CMBR by Planck Collaboration et al. (2016) have estimated the baryonic density in the early universe as $\omega_b = 0.0486 \pm 0.0010$. Similarly, Persic & Salucci (1992) estimated the baryonic density of the current universe in galaxies, groups, and clusters by integrating the baryonic mass-to-light ratio in these systems and found that the majority of the baryonic mass is unaccounted for in the present universe and got a value of $\omega_b = 0.003$. However, their calculation did not take into account the

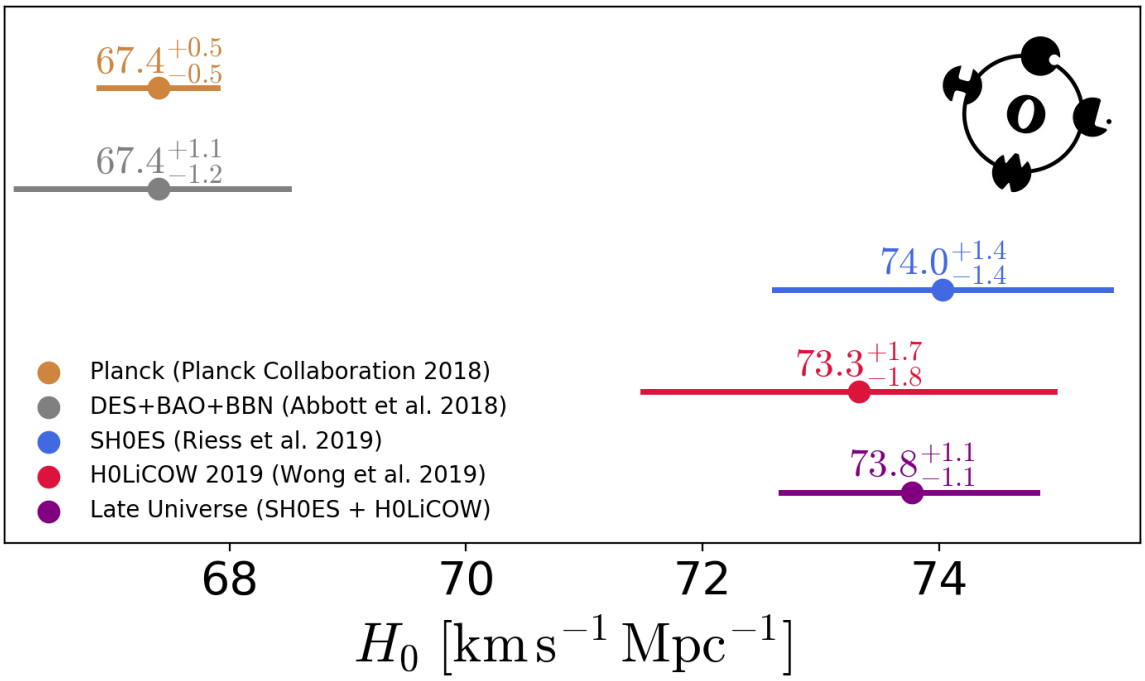


FIGURE 1.6: Comparison of the Hubble constant (H_0) with different probes within a ΛCDM universe. The orange line shows the measurement from Planck Collaboration et al. (2020). The grey line shows the combined estimate of H_0 by clustering, BAO, and BBN (Abbott et al. 2018b). The blue line displays the calculation of Riess et al. (2019) and the red line shows the estimate with SGLs by Wong et al. (2020). Image credits: Wong et al. (2020)

baryons that may reside in the intergalactic medium. Revised estimates of the baryonic mass density predict that 10% of the baryons reside in galaxies, while $50-60\%$ of the baryons reside in the intergalactic medium, and $\sim 30\%$ of the baryons are missing (Shull et al. 2012). The distribution of the baryons in the present day universe is shown in Fig. 1.7.

In summary, the baryonic density estimated for the early universe using CMBR and the baryonic mass density estimated for the present-day universe by counting the mass in the galaxies are not in agreement with each other. It is speculated that some fraction of the missing baryons must be reading in the very faint galaxies that have not been observed yet in the previous photometric surveys (Impey & Bothun 1997; O'Neil 1997), and the majority of the missing baryons reside towards the filament axes (Tuominen et al. 2021).

Small scale problems

The predictions of the ΛCDM cosmology using N-body simulations are in good agreement with the observations on scales ranging from 15000 Mpc to 1 Mpc (Schaye et al. 2015; Vogelsberger et al. 2014). However, examining the universe on smaller scales (<1Mpc) and in the realm of structures with masses less than $10^{11} M_{\odot}$, discrepancies from simulations and observations start arising, which are generally referred to as small-scale problems. It is still debated whether the small-scale problem arises because the ΛCDM model is not the correct description of the universe or due to the limited particle mass resolution in large cosmological simulations. Some of the small-scale problems are described below, and for a detailed discussion on the small-scale problems, please see Bullock & Boylan-Kolchin (2017a) or Perivolaropoulos & Skara (2022).

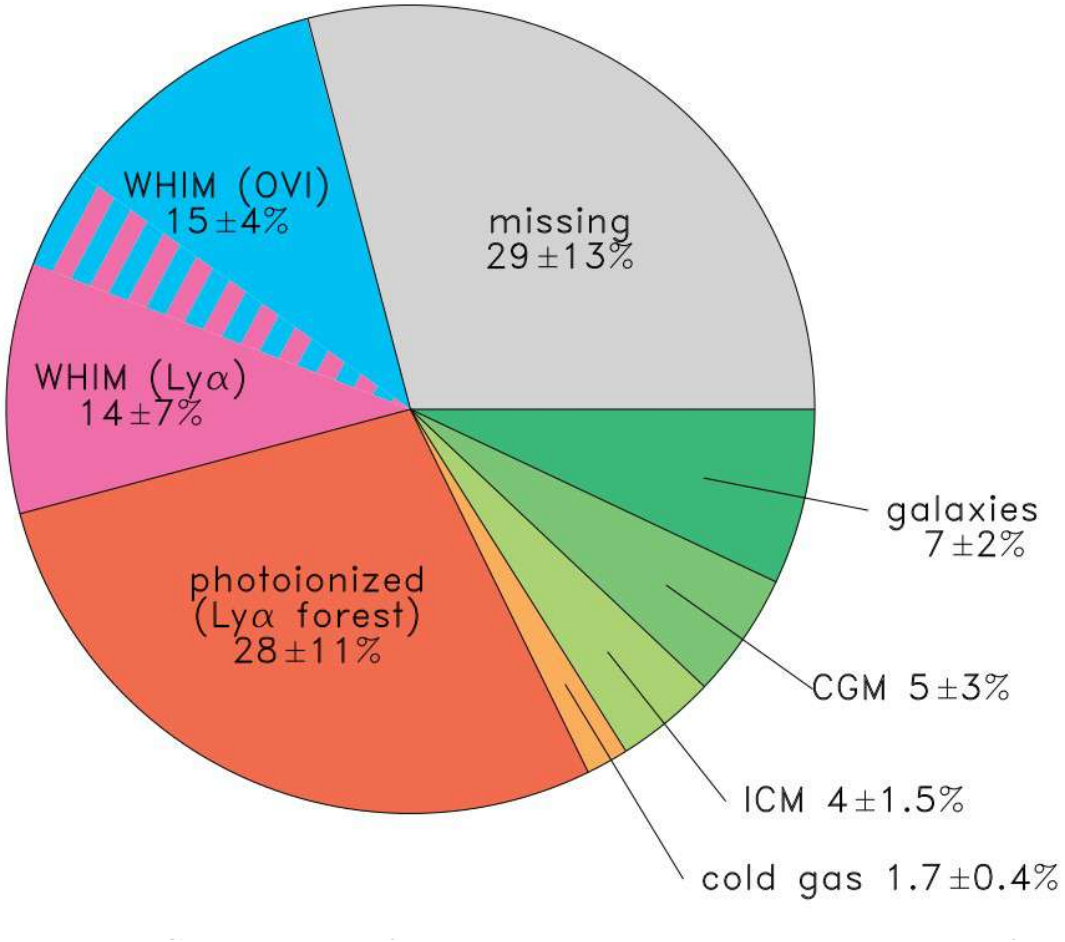


FIGURE 1.7: A pie chart showing the distribution of baryons in the universe at present. CGM represents the circum-galactic medium, ICM stands for the intracluster medium and WHIM stands for warm-hot intergalactic medium. Image credits; Shull et al. (2012).

Simulations based on the ΛCDM model of the universe predict that there will be $\sim$ 1000 subhalos capable of hosting galaxies in the local group. Typically, satellite galaxies hosted by subhaloes will have a mass of less than $10^{10} M_{\odot}$ and are considered dwarf satellite galaxies. However, we have only found 50 of these satellite galaxies in the local group (Mateo 1998; Drlica-Wagner et al. 2015), which is much less than what we expected from the simulations (Springel et al. 2008; Griffen et al. 2016). However, it should be noted that the simulations do predict the number of Milky Way-sized galaxies in agreement with observations. The discrepancy arises only for dwarf satellite galaxies, and this problem is known as the missing satellite problem. One possible solution for the missing satellites would be that these galaxies might not have been observed yet. If galaxy formation becomes less efficient at smaller halo masses, the smallest dark matter halos may not have produced enough stars to be detectable by current observational methods, which is supported by detections of ultra-faint dwarf galaxies (Simon & Geha 2007).

Another challenge faced by the ΛCDM model is known as the core-cusp problem. N-body simulations of the ΛCDM model of the universe predict that the density profile of the dark matter halos should increase steeply in the inner regions of the halo (de Blok 2010; Navarro et al. 2010). However, for galaxies, the observed rotation curves point to a flat central dark matter profile. One other problem faced by the ΛCDM model is the satellite plane, in which the observed satellite galaxies are distributed in a plane rather

than randomly, as predicted from simulations. Similarly, ΛCDM model fails to explain another problem known as the too-big-to-fail problem, where some of the massive subhalos around the Milky Way are too massive to not form satellite galaxies, yet they have failed to form satellite galaxies (Boylan-Kolchin et al. 2011). These subhalos were too large to fail to form galaxies, and they were later discovered to be true for the Andromeda galaxy as well (Tollerud et al. 2014). One possible reason for these problems might be the nature of the CDM and other dark matter candidates such as warm dark matter or fuzzy dark matter could resolve these problems (McGaugh et al. 2007). Please refer to Tulin & Yu (2018) for a detailed discussion of the small-scale problems in the ΛCDM model and how different dark matter profiles try to resolve these problems.

1.4 Laboratories to resolve the tensions of ΛCDM model

In the previous chapter, I have described only some of the challenges faced by the ΛCDM model. For a detailed discussion on all the challenges faced by the ΛCDM model, I encourage the readers to go through Famaey & McGaugh (2013); Bull et al. (2016); Popolo & Le Delliou (2017); Di Valentino et al. (2021); Perivolaropoulos & Skara (2022). Numerous proposals have been suggested to address the challenges encountered by the ΛCDM model. Many of these proposals involve introducing new forms of dark matter such as warm dark matter (Viel et al. 2013; Dekker et al. 2022) or fuzzy dark matter (Lee 2021) and dark energy such as quintessence models (Tsujikawa 2013). Others argue that modifying the theory of gravitation could explain observations without the need for dark matter or dark energy. These attempts are collectively termed modified theories of gravity (Clifton et al. 2012; Joyce et al. 2016; Shankaranarayanan & Johnson 2022). On the other hand, rather than attributing the issues to the ΛCDM model, some of the proposed solutions argue that improving our understanding of the intricate processes involved in galaxy formation and evolution will resolve the challenges faced by the ΛCDM model.

In a scenario where multiple solutions are available, the most effective approach is to test the solutions and check their consistency rigorously. Hence, to test these solutions we need to find galaxy systems which will have falsifiable predictions. In my thesis, I concentrate on the detection and analysis of two distinct types of galaxy systems: strong gravitational lenses (SGLs) and low-surface-brightness galaxies (LSBGs), which have the potential to test these proposed solutions. With the upcoming large-scale surveys such as the LSST (Ivezić et al. 2019) and Euclid (Euclid Collaboration et al. 2022b), an unprecedented number of these systems will be discovered, giving us enough statistics to resolve the tensions within the framework of the ΛCDM model or to test the new solutions.

1.4.1 Strong gravitational lenses

Strong gravitational lensing occurs as a consequence of general relativity, where the gravitational field of a massive foreground object bends and distorts the light from a background object. This distortion of spacetime around the massive foreground object leads to the formation of multiple paths for the light from the background object to traverse, resulting in the creation of multiple images of the background source. Einstein (1936) showed that the light would be distorted in the presence of a massive object such as a star, but Einstein himself did not believe that distortion could be observed as it was very small ($\sim$ 0.001 arcsec). It was Zwicky (1937) who noticed that this effect could be observed for galaxies since their masses are very high, and later it was confirmed with the detection of a strongly lensed quasar by Walsh et al. (1979).

If the strongly lensed source is a quasar or a supernova explosion which has luminosity fluctuations then the space-time distortion caused by a foreground object will result in multiple images of the source with time delays in its luminosity variations. One could show that this time delay (Δt) is inversely proportional to the Hubble constant (H_0) (Treu & Marshall 2016). Hence, one can estimate the Hubble constant independently of other cosmological probes such as the CMBR and SN-Ia by precisely measuring the time delay in lensing systems. Recently, Wong et al. (2020) estimated the Hubble constant as $73.7^{+1.7}_{-1.8}$ with 2.4% precision by measuring the time delay of 6 quasars. The six quasars used by Wong et al. (2020) are shown in Fig. 1.8.

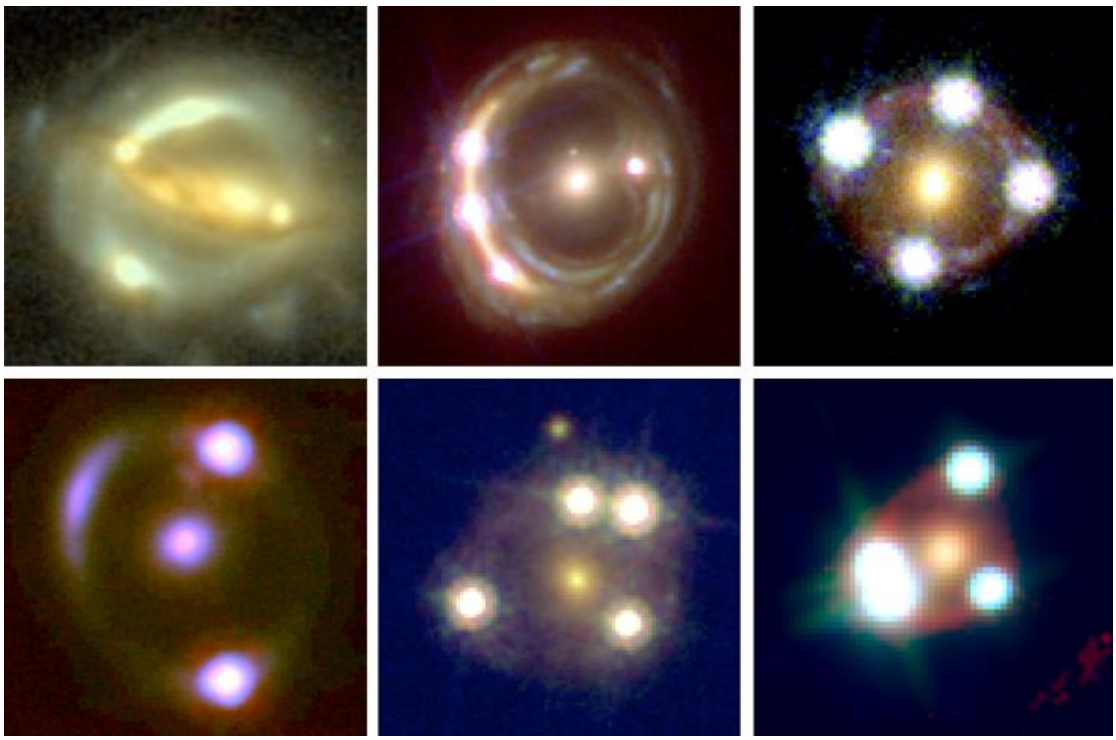


FIGURE 1.8: The images of the six quasars used by Wong et al. (2020) to estimate the Hubble constant. The centre of each image shows the lensing galaxy and the surrounding images are formed by the multiple images of the source. Image credit: Wong et al. (2020)

On the other hand, if the foreground and the background objects are galaxies, then the system is known as a galaxy-galaxy strong lens. In a galaxy-galaxy strong lensing system, if the source galaxy, lens galaxy, and observer are perfectly aligned in a straight line, a ring-like structure appears around the lensing galaxy. This ring structure, known as the Einstein ring, is the distorted image of the source galaxy, and the angular radius of this ring is referred to as the Einstein radius. The formation of an Einstein ring by the deflection of a massive galaxy is shown in Fig. 1.9. By definition, the mass (M_{lens}) inside the Einstein radius (θ_E) is given by the relation (Koopmans 2006)

$$M_{lens} = \frac{c^2}{4G}\frac{D_s D_l}{D_{ls}}\theta_E^2. \tag{1.8}$$

If we assume that a simple singular isothermal sphere (SIS) describes the density profile of the matter (baryonic+dark matter) in the lensing galaxy as:

$$\rho(r) = \frac{\sigma^2}{2\pi G r^2}, \tag{1.9}$$

where σ is the observed stellar velocity dispersion of the lensing galaxy. Then the Einstein radius (θ_E), which represents the deflection by the lensing galaxy, can be represented as:

$$\theta_E = 4\pi \frac{D_{ls}}{D_s} \frac{\sigma^2}{c^2}. \tag{1.10}$$

Here, D_s is the angular distance to the source, and D_{ls} is the angular distance between the lens and the source.

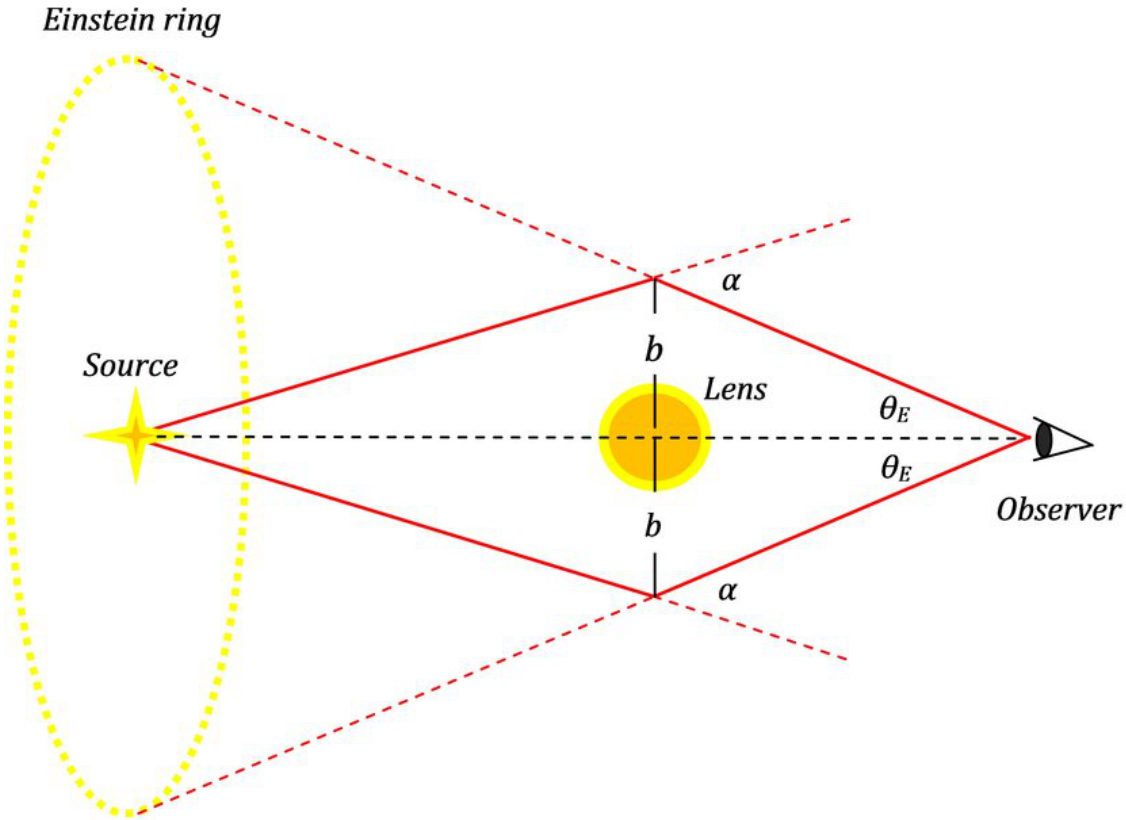


FIGURE 1.9: Formation of an Einstein ring by strong gravitational lensing.
Image credits: Pinochet & Van Sint Jan (2018).

However, the SIS model assumes a relatively simple density profile, which could be generalised by assuming a spherical power law mass distribution $\rho = r^{-\gamma}$. Where γ is known as the mass density power-law index and $\gamma = 2$ describes an SIS profile (Koopmans 2006; Koopmans et al. 2006). If we can obtain the velocity dispersion (σ_{ap}) inside an aperture of angular size θ_{ap}, then the dynamical mass inside the aperture could be expressed as

$$M_{dyn} = \frac{\pi}{G} \sigma_{ap}^2 D_l \theta_E \left(\frac{\theta_E}{\theta_{ap}}\right)^{2-\gamma} f(\gamma) \tag{1.11}$$

where

$$f(\gamma) = -\frac{1}{\sqrt{\pi}} \frac{(5-2\gamma)(1-\gamma)}{3-\gamma} \frac{\Gamma(\gamma-1)}{\Gamma(\gamma-3/2)} \left(\frac{\Gamma(\gamma/2-1/2)}{\Gamma(\gamma/2)}\right)^2. \tag{1.12}$$

Hence, galaxy-galaxy strong lenses can be used to estimate the mass of the lensing galaxy and it is one of the few observational probes capable of estimating the total mass (baryonic + dark matter). In addition, the galaxy-galaxy lenses can be used to constrain the cosmological models (Biesiada et al. 2010; Cao et al. 2015; Li et al. 2024), and study how the mass profile of the lensing galaxies evolve as a function of the redshift (Chen et al. 2019).

Observing strong gravitation lenses

As promising as it sounds, the current number of confirmed SGL systems (<250) is insufficient to constrain cosmological models precisely. In addition, we also need the velocity dispersion measured for the SGL system, which is not readily available. However, we are expected to observe and obtain spectra of around 10^5 SGL systems with large-scale photometric surveys such as LSST and Euclid (Collett & Auger 2014; Verma et al. 2019) and upcoming spectroscopic surveys such as 4MOST (de Jong et al. 2022). In this scenario, the first course of action should be to find all the SGL systems from the photometric surveys efficiently. However, SGL systems are really rare, and to find these 10^5 SGL systems, one has to analyse around 10^9 astronomical objects, which is a time-consuming task if we rely on non-automated or any other traditional methods. Hence, SGL systems could be used as an independent probe to constrain cosmological models in the era of LSST and Euclid but we need better methodologies to complete the required data. In my thesis, I take the first step in this direction by focusing on the detection of SGL systems from large-scale surveys using simulated datasets.

1.4.2 Low surface brightness galaxies

Our limited knowledge of the actual galaxy population has profound implications for nearly every aspect of observational cosmology and galaxy evolution. Our understanding of stellar populations, kinematics, dark matter content, star formation history, and large-scale clustering of galaxies beyond the Milky Way is predominantly based on studies of high surface brightness objects. For instance, Freeman (1970) compiled a sample of 36 disk galaxies with surface brightness known at that time and postulated that all the disk galaxies have a central surface brightness, $\mu_{0,B} = 21.65 \pm 0.30$ mag arcsec^{-2}. This hypothesis was known as the Freeman's law. Disney (1976) was the first to argue that Freeman's law might be arising from the selection bias and not a true characteristic of the disk galaxy populations. Disney (1976) pointed out that the sky surface brightness of the photographic plates used by Freeman (1970) is close to $\mu_{0,B} = 21.65$ mag arcsec^{-2}. Since the surface brightness of the sky background can not be removed from photographic plates unlike the CCD cameras, the galaxies fainter than the night sky were not included in the sample used by Freeman (1970) causing a biased conclusion.

The low surface brightness (LSB) regime is expected to host the majority of the galaxy population, which is supported by both simulations (e.g. Martin et al. 2019; Di Cintio et al. 2019; Jackson et al. 2021; Pérez-Montaño et al. 2022) and observations (e.g. Dalcanton et al. 1997; O'Neil & Bothun 2000). Within the LSB regime, galaxies that are fainter than the night sky are generally called low-surface-brightness galaxies (LSBGs). Since the LSBGs are poorly represented in the current galaxy catalogues, our understanding of the complete range of the galaxy population is limited. Consequently, many contradictions between theory and observation can be attributed to the LSB/dwarf regime, such as the missing satellite galaxies (Griffen et al. 2016) and the core-cusp problem (Navarro et al. 1996). Hence, the LSBG galaxies could be used as a testing ground to resolve the small-scale problems in the ΛCDM cosmology. In addition, it is speculated that some fraction of

the missing baryons must be residing in the LSB regime (Impey & Bothun 1997). In addition to LSBGs, one more candidate from the LSB regime for finding the missing baryons is known as the intra-cluster light (ICL). During interactions between galaxies in clusters, individual stars are often stripped from their parent galaxies, causing them to drift freely within the gravitational potential of the cluster (Zibetti 2008; Montes 2019). These stars give rise to the ICL in clusters.

Currently, there is no exact definition for an LSBG and in literature, the threshold values of $\mu_0(B)$ for classifying a galaxy as LSBG vary from $\mu_0(B) \geq 23.0$ mag arcsec^{-2} (Bothun et al. 1997) to $\mu_0(B) \geq 22.0$ mag arcsec^{-2} (Burkholder et al. 2001). The conventional Hubble classification fails to differentiate between LSBGs and regular galaxies (McGaugh et al. 1995), often referred to as high surface brightness galaxies HSBGs. LSBGs encompass various sub-classes distinguished by their physical size, surface brightness, and gas content. For example, ultra-diffuse galaxies (UDGs) are extended LSBGs with effective radii $R_{\rm eff} > 1.5$ kpc and central surface brightness $\mu_{g,0} < 24$ mag arcsec^{-2} (van Dokkum et al. 2015a). Another sub-class is giant LSBGs (GLSBGs), which are extremely gas-rich faint galaxies ($M_{\rm HI} > 10^{10} M_{\odot}$) that are also extended (Bothun et al. 1987; Sprayberry et al. 1995; Saburova et al. 2023). Almost dark galaxies (ADGs) form yet another sub-class of LSBGs, characterised by their extremely faint optical emissions, making them challenging to detect in wide-field surveys like the SDSS (Janowiecki et al. 2015; Leisman et al. 2017; Xu et al. 2023; Montes et al. 2024). The formation and evolution of these extreme sub-classes of LSBGs—UDGs, GLSBGs, and ADGs—are subjects of ongoing debate and serve as robust platforms to test models of galaxy evolution and cosmology (Amorisco & Loeb 2016a; Di Cintio et al. 2017; Saburova et al. 2021; Benavides et al. 2023; Laudato & Salzano 2023; Montes et al. 2024).

One major obstacle in analysing the physical properties of the LSBGs is the difficulty in obtaining their spectroscopic data and thereby their distance due to their characteristic faintness (Kadowaki et al. 2021). One solution to this problem is to estimate the photometric redshift (photo-z) of the LSBGs and use the photo-z to estimate the physical properties of the LSBGs (Junais et al. 2023). Another alternative solution is to assume that the LSBGs within the virial radius of a cluster or group are members of it and use the redshift of the cluster to estimate the physical properties of the LSBGs, such as size and stellar mass. In my thesis, I use the sample of LSBGs within the virial radius of clusters to estimate the physical size and subsequently study the population of LSBGs and UDGs in these clusters.

In the local universe, unlike luminosity, the surface brightness of an extended source is independent of the distance between the source and the observer. The reason for this lies in the fact that the flux from an extended source falls as the inverse square of its distance, $(1/d^2)$, whereas the area within a 1 arcsec^2 region increases as d^2, effectively cancelling out each other. However, for sources at high redshifts, one should also take into account the cosmological dimming factor, where $\mu \propto (1+z)^{-4}$, which makes the sources at high redshifts appear fainter than their true surface brightness. Hence the lack of redshift of the LSBG samples will sometimes result in minor contamination from high-z sources.

One compelling reason to study LSBGs is their significance as the dominant population among galaxies. For example, Martin et al. (2019) estimated the number density of galaxies as a function of their effective surface brightness in the r-band using the Horizon-AGN cosmological simulations across various environments (field, group, and cluster), which is shown in Fig. 1.10. In Fig. 1.10, galaxies are categorized into HSBGs, classical LSBGs, and UDGs based on their surface brightness. It is clear from Fig. 1.10 that the majority of the galaxy population resides in the faint regime, irrespective of the environment. Therefore, without understanding the LSBG population, our models of galaxy evolution will remain incomplete, and understanding the formation and evolution of LSBGs will help us to alleviate observational tensions on the galaxy scale.

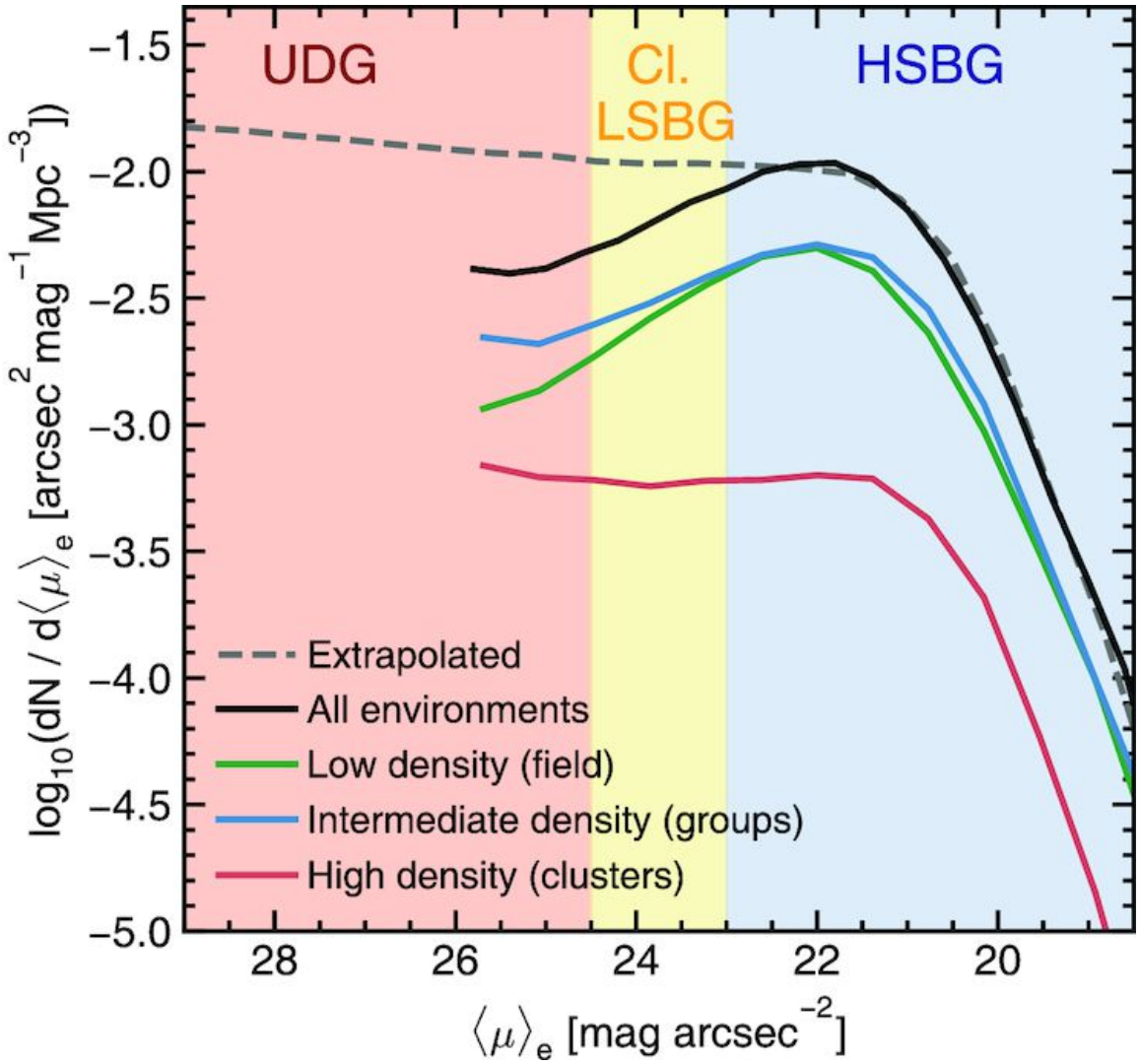


FIGURE 1.10: The number density of galaxies as a function of their r-band effective surface brightness at $z = 0$. The number density in the field, groups, and clusters is represented by green, blue and red colour curves. The black curve represents the total number density, irrespective of the environment. The blue, yellow, and red-shaded regions represent the HSBGs, classical LSBGs, and UDG populations, respectively. The dashed line indicates the extrapolated surface-brightness function with stellar mass down to $10^7 M_\odot$. Image credits: Martin et al. (2019).

Ultra-diffuse galaxies

As mentioned earlier, ultra-diffuse galaxies (UDGs) represent a subclass of LSBGs characterised by their considerable size, comparable to that of Milky Way-like galaxies, yet exhibiting very faint luminosities akin to dwarf galaxies. van Dokkum et al. (2015a) was the first to detect UDGs in significant numbers (47) in the Coma cluster, and later searches for UDGs found that Coma alone hosts around 10^3 UDGs (Koda et al. 2015; Yagi et al. 2016; Bautista et al. 2023). Similar searches in the other clusters have also found that UDGs are abundant in clusters (van der Burg et al. 2016) such as Virgo (Lim et al. 2020) and Fornax (Venhola et al. 2022). Later studies have shown that the abundance of UDGs scales linearly with host halo mass (van der Burg et al. 2016; Mancera Piña et al. 2018), and massive clusters can host thousands of UDGs (Janssens et al. 2019). It should also be noted that the

UDGs are not exclusively found in clusters, as they can be found in groups (Cohen et al. 2018; Marleau et al. 2021) as well as in field environments (Prole et al. 2019).

The term 'UDG' was coined by van Dokkum et al. (2015a); however, such galaxies were identified in several earlier studies in the literature (Sandage & Binggeli 1984; McGaugh & Bothun 1994; Dalcanton et al. 1997; Conselice et al. 2003a). van Dokkum et al. (2015b) initially suggested that UDGs are failed Milky Way-like galaxies hosted by dark matter halos similar to our Milky Way ($\sim 10^{12} M\odot$). In contrast, Amorisco & Loeb (2016b) argued that UDGs are not failed galaxies but rather the tail end of the dwarf galaxy population, which has dark matter halos with high angular momentum. One other formation scenario of UDGs suggested is the inflation of a dwarf-like progenitor because of internal processes such as stellar feedback (Di Cintio et al. 2017; Chan et al. 2018) or external processes such as tidal interaction (Carleton et al. 2019; Jones et al. 2021) or mergers (Conselice 2018) or ram pressure stripping (Junais et al. 2022). Observations indicate that UDGs in the field are richer in HI gas, tend to be bluer and have an irregular morphology in comparison to the UDGs in the clusters indicating that the environment can also be a factor in the formation and characterization of UDGs (Barbosa et al. 2020; Kadowaki et al. 2021)..

UDGs also exhibit extremes in dark matter content, with examples like DF44 containing 99.99% dark matter (van Dokkum et al. 2016), while NGC 1052-DF2 is nearly devoid of dark matter (van Dokkum et al. 2018). The colour image of the NGC 1052-DF2 galaxy is shown in Fig. 1.11. The dynamics and formation of these galaxies are still debated and these galaxies could be used as a testing ground to test our models on dark matter and galaxy evolution. For instance, the formation scenario of UDGs like NGC 1052-DF2 or DF44 has not been accurately explained by CDM-based models (van Dokkum et al. 2016, 2018). These discrepancies have led to investigations into alternative dark matter theories and modified gravity theories to better understand the origins of these galaxies (Haghi et al. 2019; Laudato & Salzano 2023; Mancera Piña et al. 2024; Golini et al. 2024).

FIGURE 1.11: Image of NGC 1052-DF2 taken with the Hubble space telescope. Image credits: van Dokkum et al. (2018)

As I mentioned earlier, getting the spectroscopy of the structures in the LSB regime is a hard and time-consuming task due to their characteristic faintness. This applies to the case of UDGs as well, limiting our knowledge of the physical properties of the UDGs. As of now, the spectroscopic data of only less than 50 UDGs has been observed and used to confirm their status as UDGs. Hence, it should be pointed out that all the UDGs identified from the clusters, groups, and fields can only be considered UDG candidates and confirmed as UDGs after knowing their physical size by measuring their redshift. Since measuring the redshifts of the UDG candidates studied in this thesis is beyond the scope of this work, for simplicity, I prefer to refer to the UDG candidates as UDGs.

Observing the low surface brightness universe

Due to their characteristic faintness, the detection of structures within the LSB regime poses significant observational challenges and has limited our exploration of these systems. In recent years, advancements in digital imaging have enhanced our ability to detect LSBGs, despite the persistent challenges. The upcoming large-scale surveys such as LSST and Euclid are expected to find more than 10^5 LSBGs (Thuruthipilly et al. 2024b) and going to revolutionise our understanding of galaxy evolution. However, the LSB regime is plagued by artefacts and often needs a visual inspection to separate LSBGs and artefacts (Greco et al. 2018; Tanoglidis et al. 2021b) which would not be a practical solution in the era of LSST and Euclid. In my work, I focus on the detection and the analysis of LSBGs using machine learning (ML) to better prepare for the LSST and Euclid era. For this, in this thesis, I develop deep learning-based models to identify LSBGs and UDGs from large-scale surveys such as DES and also on deep datasets such as those obtained from the Hyper Supriem Cam (HSC). I follow the definitions of Greco et al. (2018) and Tanoglidis et al. (2021b) to classify all the galaxies that have a mean surface brightness ($\bar{\mu}_g$) within the effective radius greater than 24.2 mag arcsec^{-2} and a half-light radius ($r_{1/2,g}$) greater than 2.5 arcsec in the $g-$band as LSBGs.

1.5 Astronomy in the era of big data

In the previous sections, I briefly gave an overview of the ΛCDM model of cosmology and discussed the observable characteristics of galaxies that probe the universe. I also highlighted some of the challenges encountered by the ΛCDM model and how SGLs and LSBGs can be utilized to constrain as well as test the alternatives to the ΛCDM model. Both SGLs and LSBGs are situated at the extreme ends of the galaxy population in terms of mass as well as luminosity. Similarly, the SGL systems are very rare whereas the LSBGs constitute the most dominant fraction of the galaxy population. However, one common feature of both of these two systems is the difficulty in finding them. For instance, the number of confirmed SGL systems is less than 300, and the previous searches for LSBGs from large-scale surveys had to visually inspect the samples to remove the artefacts, which were around $\sim$ 50% of the sample size (Greco et al. 2018; Tanoglidis et al. 2021b).

In the coming decade, the amount of data generated in astronomy will exponentially increase, as shown in Fig. 1.12. Consequently, more efficient methodologies will be necessary to analyze these large volumes of data to find SGLs and LSBGs, as well as to manage astronomical data in general. In this scenario, the regime of ML automatically becomes the natural choice, as the regime with big datasets is where the ML models excel. In addition, as the volume of data grows, ML models are expected to improve and become even more efficient, surpassing human capabilities in data analysis. For instance, in a recent study comparing the efficiency of different methodologies in finding SGLs, it was found that ML

models performed better than human experts (Metcalf et al. 2019). Hence, in the coming age of data, ML models could be used to identify complex patterns more efficiently in astronomy. In my thesis, one of my major focuses is to develop state-of-the-art ML-based methods for the identification of SGLs and LSBGs, which I will explain in the upcoming chapter.

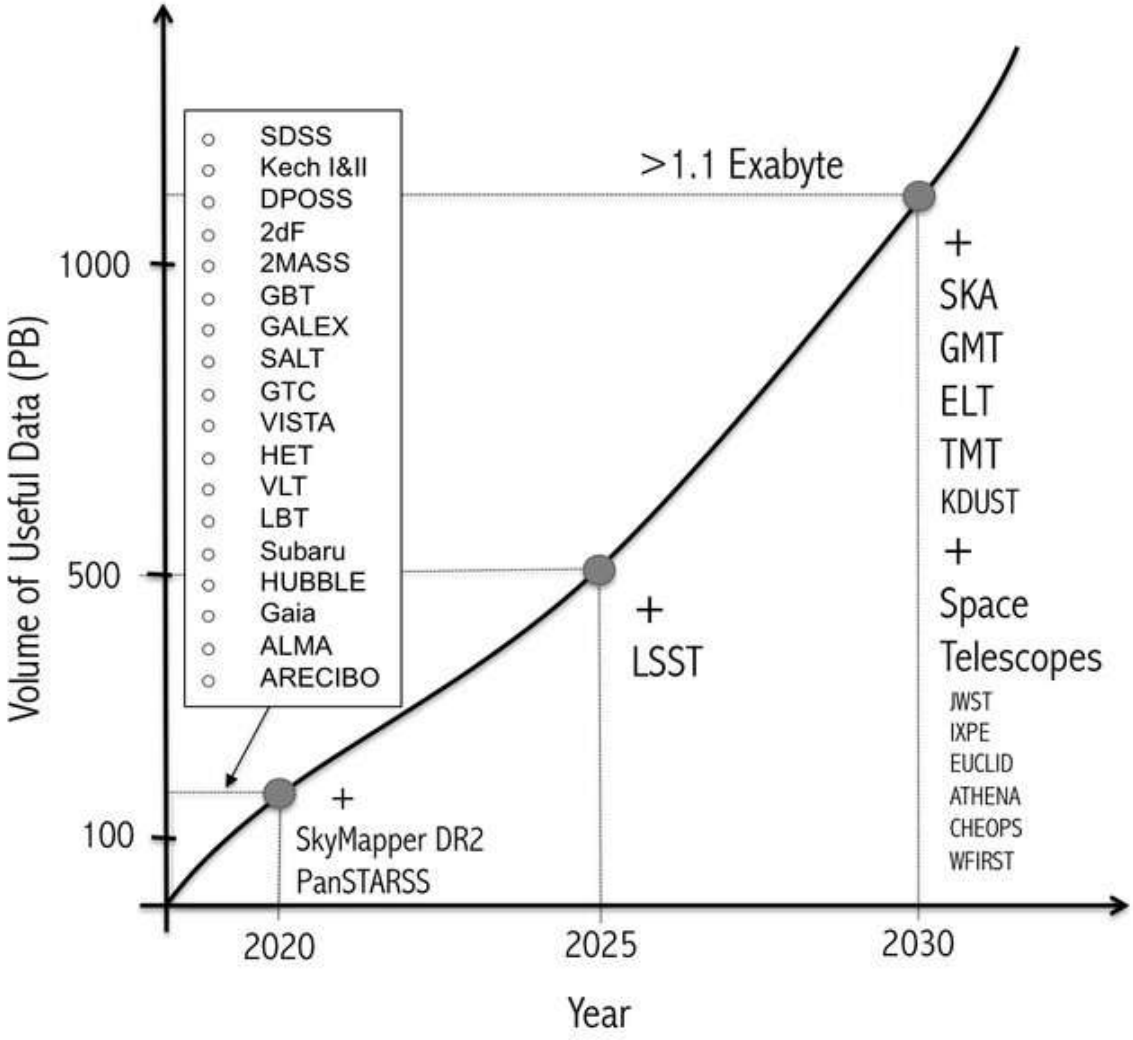


FIGURE 1.12: The evolution of the volume of data generated from major astronomical surveys and projects over the years. Image credits: Rosa (2020).

2

Introduction to machine learning

The regime of machine learning (ML) has exponentially developed in the last decade, thanks to ever-growing datasets and the introduction of new concepts and ideas in ML. Unlike conventional computer programmes, which rely on predefined logic to generate outputs based on inputs, ML operates on a different premise. In ML, the system autonomously learns the relationship between inputs and outputs, bypassing the need for explicitly programmed logic. The advantage of this is that the systems can learn to identify complex patterns similar to those of humans and sometimes work more efficiently than humans.

2.1 Machine Learning Taxonomy

Any algorithm that improves its performance with iterations could be considered a machine learning algorithm. A simple example of an ML algorithm is linear regression, in which we try to map the outputs (y_i) to the inputs (x_i) under the assumption that they follow the relation $y_i = mx_i + c$. From different combinations of m and c, the optimum value of m and c that best describes the relation between x and y can be found by minimising the error between the output y_i and the value calculated from $mx_i + c$. This process is commonly called curve fitting. One could say ML algorithms are sophisticated curve-fitting algorithms with parameters as high as 10^9. Generally, all the ML algorithms try to minimise the error between the predictions and the true outputs for a given input during their training phase. The error function that the ML tries to use during its training phase is generally known as the cost function.

ML algorithms can be classified based on the need for human supervision into two types: **supervised** and **unsupervised** learning. In supervised ML, the algorithm is trained using labelled data, allowing it to differentiate between various classes or predict outputs based on provided input-output examples. One of the drawbacks of supervised ML is the need for a labelled dataset as a training sample, and if there are any errors or problems in the training dataset, it would also be reflected in the performance of the ML model. Unsupervised ML algorithms tackle this problem by removing the need for a training set. Unsupervised ML algorithms could be used to cluster data points with common features and give labels to each cluster to achieve classifications similar to supervised ML algorithms. Other examples of the applications of unsupervised ML algorithms are dimensionality reduction and principle component analysis. However, unsupervised learning models have their own set of challenges, such as the lack of ground truth, ambiguity in clustering, and the curse of dimensionality. For a detailed discussion on machine learning, please refer to Sammut & Webb (2010) or Ivezić et al. (2014).

2.2 Deep learning

Deep learning is a subset of ML algorithms that use artificial neural networks (ANNs) or similar networks inspired by the human nervous system. Deep learning-based models perform exceptionally well in analysing complex patterns, such as in images, compared to other ML methods. Generally, creating a deep learning model requires more data and longer training times compared to other machine learning models. However, the results from deep learning models have consistently demonstrated superior performance (Janiesch et al. 2021). One fascinating feature of the deep learning model is its capacity to generalise the data and go beyond the training data to apply learned patterns to new, unseen data (Bay & Yearick 2024). In contrast, traditional models like polynomial regression are limited to interpolating within the provided data range and cannot extrapolate beyond its boundaries.

In this chapter, I will introduce the basics of deep learning and describe the different deep learning models.

2.2.1 Perceptron & artificial neuron

The fundamental building block of ANNs is known as the perceptron, which tries to mimic the action of a neuron in the human nervous system. Mcculloch & Pitts (1943) introduced the perceptrons as binary classifiers that can distinguish between linear classes. For a set of inputs (x_i), the perceptron classifies them into different classes based on the value of the function: $f(x) = \theta(\vec{w}.\vec{x} + b)$, where θ is the Heaviside step-function (outputs 0 or 1 based on a threshold), $\vec{x}$ represents the input vector, $\vec{w}$ is the vector of real-valued weights, and b is a real number known as bias. The bias term allows the model to adjust the output independently of the input values. This is especially useful when input features are zero but the true output is non-zero, ensuring the output defaults to the bias value instead of zero. By optimising the $\vec{w}$ and the bias term, one can use perceptron as a linear binary classifier.

If the decision boundary between the classes is non-linear, the perceptrons cannot be used for classification. One can overcome this limitation by using a non-linear function such as a sigmoid or tanh instead of the Heaviside step function, allowing for the classification of non-linear classes. In this scenario, the non-linear function used is known as the activation function. The modified structure of the perceptron is often called an artificial neuron because of its resemblance to the neuron. The structure of the artificial neuron and a neuron from the human nervous system is shown in Fig. 2.1 for comparison.

2.2.2 How deep learning models learn

The learning part of any DL model happens during the phase known as training. In the training phase, the DL model tries to optimise the weight vectors ($\vec{w}_i$) of the network to learn the patterns in the input data correctly and minimise the error between the prediction and the true output. Usually, in the training phase, there are five steps involved, which are listed below.

- Initialization: The weights of the neural networks are set to random values. This is important because if all the weights had the same values, then the output of each artificial neuron would be the same in a layer, making it difficult for the network to learn the features.
- Forward Propagation: The input data is propagated forward through the network by computing the output of each artificial neuron in each layer.

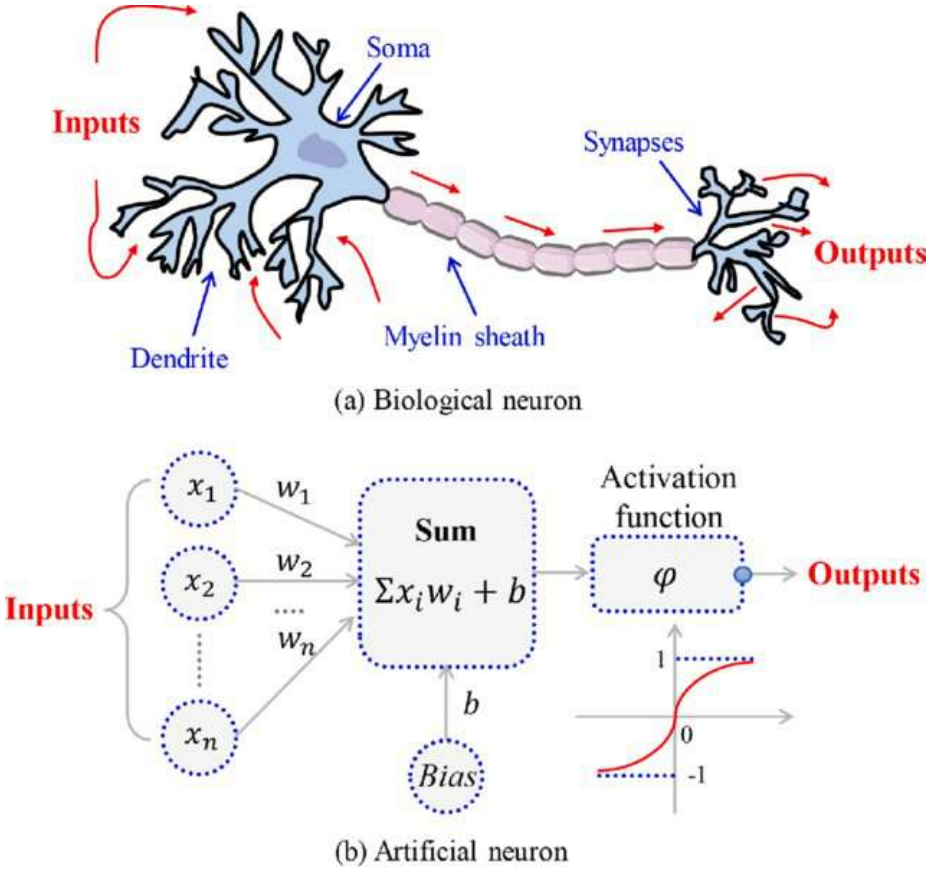


FIGURE 2.1: Comparison between a biological neuron (a) and an artificial neuron (b) with sigmoid $(1/1+e^{-x})$ as the activation function. Image taken from Wang et al. (2021)

- Loss calculation: After forward propagation, the output predicted by the final layer of the network is compared to the true output value to calculate the prediction error. This error is estimated with a predefined function known as the loss function (J).
- Backward propagation and weight update: Using the gradient of the estimated error with respect to the weight, the value of each weight is updated in such a way that it will minimize the prediction error. Mathematically, the update of an individual weight vector can be represented as

$$w_{\text{new}} = w_{\text{old}} - \alpha \frac{\partial J}{\partial w}. \tag{2.1}$$

  Here, α is known as the learning rate of a model and it decides the speed at which the model learns the patterns in the data.
- Iteration: Steps from 2–4 are repeated until the value of the loss function (also referred to as simply loss) no longer decreases with repetition and the value is small. Each iteration is also known as an epoch.

For creating a robust DL model, it is essential to divide the available data into three different sets: training, validation, and test datasets. The training data is used to train the DL model, whereas the validation data is employed to validate the performance of the model at each step during training. The difference between the training data and the validation data is that steps 1–5 are performed with the training data, whereas the validation data

is not used for this. Validation data is used to estimate the loss (validation loss) at each iteration and compare it with the training loss. If both the training loss and validation loss plateau and do not decrease after several iterations, it indicates that the DL model has successfully learned the patterns in the data. Generally, the validation loss will be less than or similar to the training loss at each epoch. If the training loss is significantly less than the validation loss, it suggests that the model has overfitted to the training data. Hence, the validation dataset is crucial for monitoring the training process of the DL model at each epoch, helping to ensure that the model generalizes well to new, unseen data. Finally, the test data is used to evaluate the performance of the model after training. In practice, during training, instead of the whole training data, small batches of training data are used to update the weights of the DL model in each epoch. Since it would be computationally expensive to load the entire training data in one epoch. One another important fact to be considered during training is that the training dataset used to train the DL model should be balanced. Otherwise, the DL model will fail to identify the patterns within the data.

2.3 Artificial neural networks

Artificial Neural Networks (ANNs) are computational models designed to recognize patterns in data, inspired by the human nervous system. The universal approximation theorem is fundamental to understanding how ANNs learn and represent complex patterns. It states that a neural network with a single hidden layer can approximate any continuous function, given sufficient artificial neurons in the hidden layer and an appropriate non-linear activation function in the artificial neurons (Hornik et al. 1989). However, in practice, it is computationally more efficient to have more interconnected layers with a few artificial neurons than to have one single hidden layer with a large number of artificial neurons.

The fundamental units of ANNs are artificial neurons which are organized into layers, and each neuron in one layer is connected to neurons in the next layer through weighted connections. The layers that intake the input vectors are known as the input vectors, and the layer that produces the final outputs from the network is known as the output layer. Generally, the layers between the input and output layers are known as the hidden layers of a DL model. An example of a 2-layer ANN (1 hidden layer and one output layer) is shown in Fig. 2.2. In this example, the ANN takes 2 features as input and passes them to the hidden layer, which consists of 5 artificial neurons. The output from this layer is passed to the output layer to predict the three outputs in the output layer.

If $\vec{X}_{i-1}$ represents the input to the i^{th} layer of the ANN, W_i denotes the weight matrix corresponding to the i^{th} layer, and $\vec{b}_i$ is the bias vector, then the output of the i^{th} layer can be computed as:

$$\vec{X}_i = \phi\left(W_i \vec{X}_{i-1} + \vec{b}_i\right). \tag{2.2}$$

Similarly, during training the weight matrix and the bias vector are updated following the relation

$$W_{i,new} = W_{i,old} - \alpha \frac{\partial J}{\partial W_{i,old}} \tag{2.3}$$

$$\vec{b}_{i,new} = \vec{b}_{i,old} - \alpha \frac{\partial J}{\partial \vec{b}_{i,old}}, \tag{2.4}$$

where α is the learning rate of the ANN. Generally in an ANN, all the artificial neurons in a layer are connected to all the artificial neurons in the previous and the next layer as similar to shown in Fig. 2.2. Because of this ANNs are sometimes also called fully connected

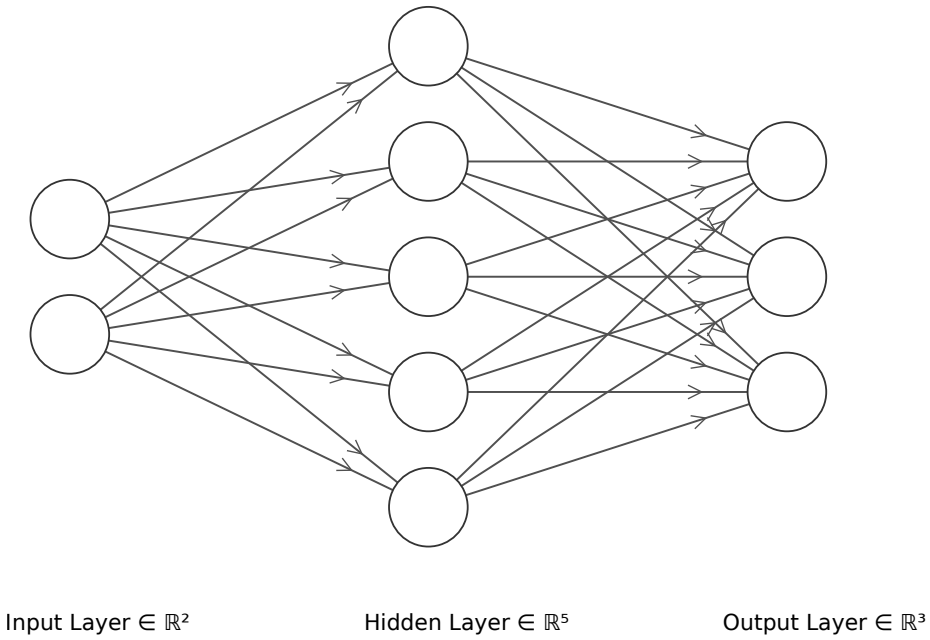


FIGURE 2.2: Schematic diagram of a 2-layer ANN. The input layer has 2 artificial neurons and the output layer has 3 artificial neurons. The hidden layers have 5 artificial neurons.

neural networks. For a detailed review on ANNs, please see Schmidgall et al. (2023).

2.4 Convolutional neural networks

The concept of using convolution neural networks (CNNs) to analyse image-like data was first proposed by Lecun et al. (1998). The CNNs were created based on the structure of the human visual cortex and how it processes spatial features. However, a breakthrough for image recognition by CNNs did not happen until Krizhevsky et al. (2012) created an architecture that won the ImageNet Large-Scale Visual Recognition Challenge 2012. Since then, CNNs have been extensively employed in various research disciplines, following the proposed architecture. One advantage of CNNs over ANNs in analyzing image data is their ability to directly process the image without the need to manually identify features and feed it to the network like ANNs.

A regular CNN can be thought of as consisting of two parts. The first part consists of convolution layers and pooling layers which extract the features of the image. The convolution layers can learn the local spatial correlation in the data, hence using multiple convolution layers will help us to detect the features in the data (Mallat 2016). The convolutional layers consist of multiple matrices typically of size 3×3 or 5×5 which are also referred to as filters or kernels. However, a CNN is restricted by the size of its kernels to collect spatial information from the data. Hence, it may lead to poor performance in some cases due to the ignorance of global information. Generally, the convolutional layers are followed by a pooling layer to downsample the size of the feature maps to reduce the computational cost. After a series of convolutional layers, all the feature maps are flattened and joined together in sequence to connect them to a series of fully connected (FC) layers. This is the second part of the CNN which resembles the structure of usual ANNs. The FC layers process the extracted features of the image data to predict the required output.

The operation of the convolutional layers is based on the kernels or the filters. Each filter in the convolution layer convolves the input data to generate the output known as

feature maps which are multiplied by a weight and passed on to the filter in the next convolutional layer. If we have k number of input features passed on to nth convolution layer in the CNN that outputs l number of output feature maps, then the output feature maps are computed based on the relation:

$$x_n^l = \phi\left(W_n^{k,l} * x_n n - 1^k + b_n^l\right) \tag{2.5}$$

Here, $W_n^{k,l}$ is the weight matrix that contains the filter of the n^{th} layer, $*$ represents the convolution operation and b_n^l represents the bias matrix of the n^{th} layer. During the training phase, the kernels are optimised to learn spatial features. For a detailed discussion on CNNs, please refer to Khan et al. (2020) and Alzubaidi et al. (2021).

2.5 Transformers

While Convolutional Neural Networks (CNNs) have been the dominant choice for analyzing image data in astronomy, the latest state-of-the-art models for image analysis are transformers. Transformers were initially introduced in natural language processing (NLP) as attention-based models (Vaswani et al. 2017). The fundamental concept behind transformer architecture is the attention mechanism, which has found broad applications across machine learning (Zhang et al. 2019; Fu et al. 2019; Parmar et al. 2019; Zhao et al. 2020; Tan et al. 2021). In NLP, attention mechanisms calculate the correlation of different positions within a single sequence to derive its representation. This concept was later adapted for computer vision, leading to state-of-the-art models for various image processing tasks such as image classification (Wortsman et al. 2022) and image segmentation (Chen et al. 2023).

Transformers in computer vision can be categorized into two types. The first type combines both CNNs and attention mechanisms. This approach incorporates the strengths of both architectures: CNNs are excellent at local feature extraction and capturing low-level details and spatial hierarchies, while attention layers excel at modelling global context and long-range dependencies. An example of this is the Detection Transformer (DETR), proposed for end-to-end object detection by Carion et al. (2020). The second type of transformer operates entirely on self-attention mechanisms without using any convolutional layers. An example is the Vision Transformer (ViT), proposed for object classification by Dosovitskiy et al. (2021). ViTs have demonstrated remarkable performance in image classification tasks, surpassing the accuracy of CNN-based models on various benchmark datasets (Dosovitskiy et al. 2021; Yu et al. 2022; Wortsman et al. 2022).

In the following subsections, I will briefly introduce the concept of self-attention which is at the core of the transformer models, and then discuss the two transformer models (DETR & ViT) mentioned above in detail. Throughout my thesis, I will use models based on these two transformer models (DETR & ViT) to identify and classify strong gravitational lenses and low surface brightness galaxies.

2.5.1 Self-attention

The introduction of attention mechanisms in machine learning has the potential to revolutionise machine learning, and it has been found particularly useful in NLP. Depending on the task at hand, various attention mechanisms can be employed. For a review of various attention mechanisms, please refer to Yang (2020); Niu et al. (2021). The idea of the attention mechanism in transformers is to dynamically focus on relevant parts of the input data, similar to how humans pay attention to important information in a conversation

while ignoring irrelevant background noise. Mathematically, attention can be defined as

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \tag{2.6}$$

where Q, K, V are vectors and $\sqrt{d_k}$ is the dimension of the vector key (K). The softmax function, by definition, is the normalised exponential function that takes an input vector of K real numbers and normalises it into a probability distribution consisting of K probabilities proportional to the exponential of the input numbers. As we compute the normalised dot product between the query (Q) and the key (K), we get a tensor (QK^T) that encodes the relative importance of the features in the key to the query (Vaswani et al. 2017).

Among various attention mechanisms, self-attention is one of the most widely used attention mechanisms for image analysis. The main idea of self-attention is to assign relative importance to the features of the input based on the input itself. For self-attention, the vectors (Q), (V), and (K) are identical. Hence multiplying the tensor (QK^T) by vector (V) results in a vector that encodes the relative importance of features inside the input vector. In other words, the central idea of self-attention is to assign relative importance to the features of the input based on the input itself. A physical interpretation of self-attention applied to feature vectors can be thought of as filtering the input features based on the correlation in the input. The structure of a multi-head attention layer is given in Fig. 2.3. It is possible to provide self-attention with more power by creating several layers and dividing the input vector into smaller parts (H, number of heads). Each attention layer is called a head, which applies self-attention to one part of the divided input.

Positional encoding

If we pass the input directly to the attention layers, the input order or the positional information is lost as transformer models are permutation invariant. So to preserve the information regarding the order of features, we use positional encoding, and the lack of positional encoding will lower the performance of a transformer model. Following the work of Vaswani et al. (2017), I use fixed positional encoding defined by the function

$$PE_{(pos,2i)} = \sin\left(\text{pos}/12800^{\frac{2\text{i}}{\text{d}_{\text{model}}}}\right), \tag{2.7}$$

$$PE_{(pos,2i+1)} = \cos\left(\text{pos}/12800^{\frac{2\text{i}}{\text{d}_{\text{model}}}}\right), \tag{2.8}$$

where pos is the position, i is the dimension, and d_{model} is the dimension of the input feature vector. Each dimension of the positional encoding corresponds to a sinusoid function. For a detailed description of positional encoding and its importance, I refer the reader to Vaswani et al. (2017); Liutkus et al. (2021); Su et al. (2024); Chen et al. (2021).

2.5.2 Detection Transformer

One class of transformer models that I created are inspired by the DEtection TRansformer (DETR) created by Facebook (Carion et al. 2020). For simplicity, we call these models DETR, but it should be kept in mind that the model presented in Carion et al. (2020) is not the same as the transformer model presented here. The structure of the DETR is simple and is similar to a CNN except for the addition of self-attention layers in between the convolutional layer and the fully connected layer. The extra component in between

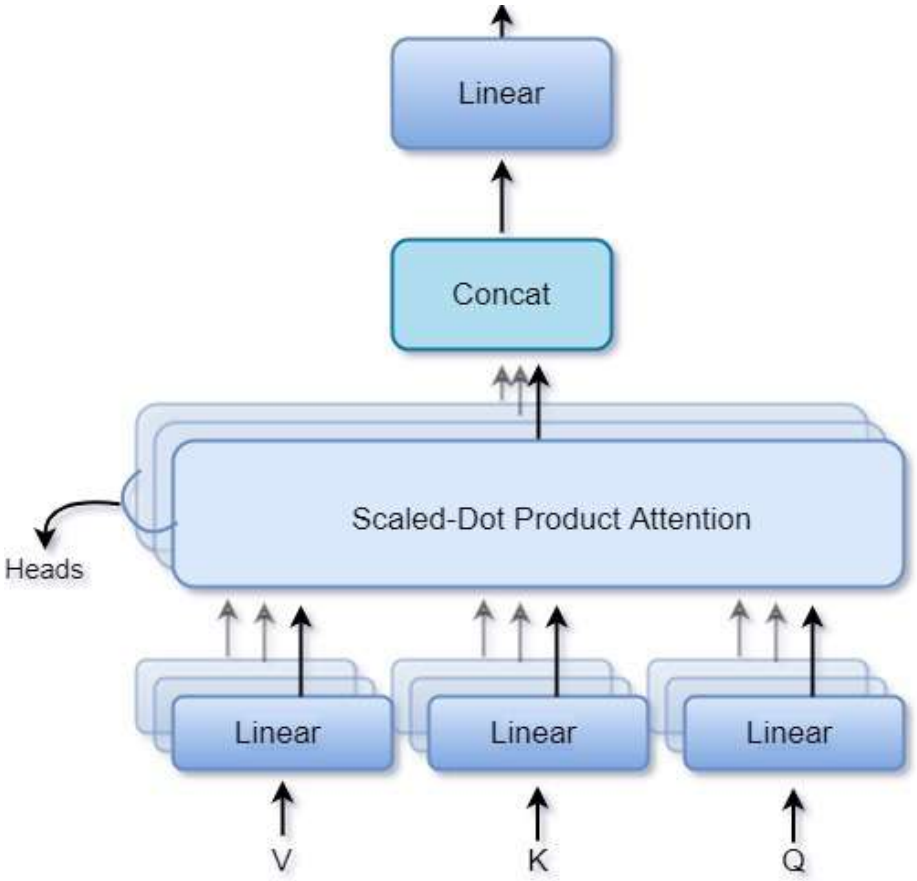


FIGURE 2.3: Scheme of the multi-head attention layer. The input vector is split into small vectors V, K and Q and passed on to the attention layer with multiple heads where the output is computed as per Eqn. 2.6. The resulting sequences from multiple heads are concatenated to form a single sequence output vector. Image taken from Thuruthipilly et al. (2022).

the convolutional layer and the fully connected layers is called an encoder layer or transformer encoder owing to its origin from NLP. The encoder layer has a standard architecture and consists of a multi-head self-attention module as shown in Fig. 2.3 followed by a feed-forward network (FFN). The structure of the DETR is shown in Fig. 2.4.

Similar to CNNs, the DETR can be considered to have three components. The first component of the architecture is a sequence of convolutional layers known as the CNN backbone to extract the features from the image. The output from the CNN backbone is a vector with dimensions H$\times$W$\times$D, where D is the number of filters in the last convolution layer. The second component which is the self-attention layers demands a sequence as input. Hence the output from the CNN backbone should be reshaped to a D$\times$HW feature map. However, when reshaping the information about the spatial positions will be lost and make the transformer architecture permutation-invariant. Hence, before passing the output from the CNN backbone to the encoder layer, we add the output of the CNN backbone with fixed positional encoding. The output from the CNN backbone added with positional encoding is then processed by the encoder layer to filter out only the relevant features extracted by the CNN. The third part of the DETR model is an FFN that is similar to the regular CNNs and that learns the features filtered by the encoder layers.

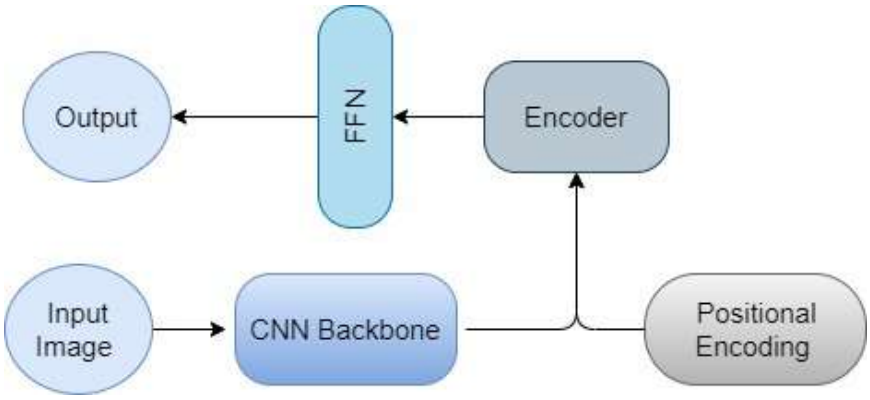


FIGURE 2.4: Scheme of the architecture of the detection transformer (DETR). Image taken from Thuruthipilly et al. (2022).

2.5.3 Vision Transformer

Another class of transformer models that I have created is based on the Vision Transformer (ViT) introduced by Google Brain (Dosovitskiy et al. 2021), which we also call ViT for simplicity. One of the main features of ViT models is that it does not use any convolutional layers to process the image, unlike DETR models. In the ViT architecture, the input image is divided into fixed-size patches, which are flattened into a sequence of 1D vectors. Since the transformers are permutation invariant, the positional embedding is added to the patch embedding before they are fed into the transformer layers. The positional embedding is typically a fixed-length vector that is added to the patch embedding, and it is learned during training along with the other model parameters. The combined 1-D sequence is then passed through a stack of transformer layers. An additional learnable (class) embedding is affixed to the input sequence, which encodes the class of the input image. This class embedding for each input is calculated by applying self-attention to positionally embedded image patches. Output from the class embedding is passed on to a multi-layer perceptron (MLP) head to predict the output class. A schematic diagram of the Vision transformer is shown in Fig. 2.5. For a detailed discussion on ViT models, please refer to Dosovitskiy et al. (2021).

2.6 Transfer learning

Generally, transfer learning in machine learning refers to the practice of re-using a pre-trained model for a new task. With transfer learning, we are trying to transfer the knowledge gained by a DL model for one task to better understand the features in another. In computer vision, deep learning models such as CNNs or Transformers typically aim to detect edges in the first layers and learn how to integrate these edges for understanding task-specific features in the latter layers. Hence, the features learned in the first few layers of a DL model are general and the weights of the initial layers can be effectively transferred from one task to another, facilitating faster and more efficient learning.

Transfer learning has been found to be very useful and successful in training large natural language models and computer vision models. Given the need for better machine learning models and techniques in the era of big data, the astronomy community has also delved into the potential of transfer learning. For instance, Ackermann et al. (2018) used a CNN model that has been trained on pictures of everyday objects (i.e. ImageNet data set, Deng et al. (2009)) to be retrained for the detection of galaxy mergers. Similarly, Wei et al.

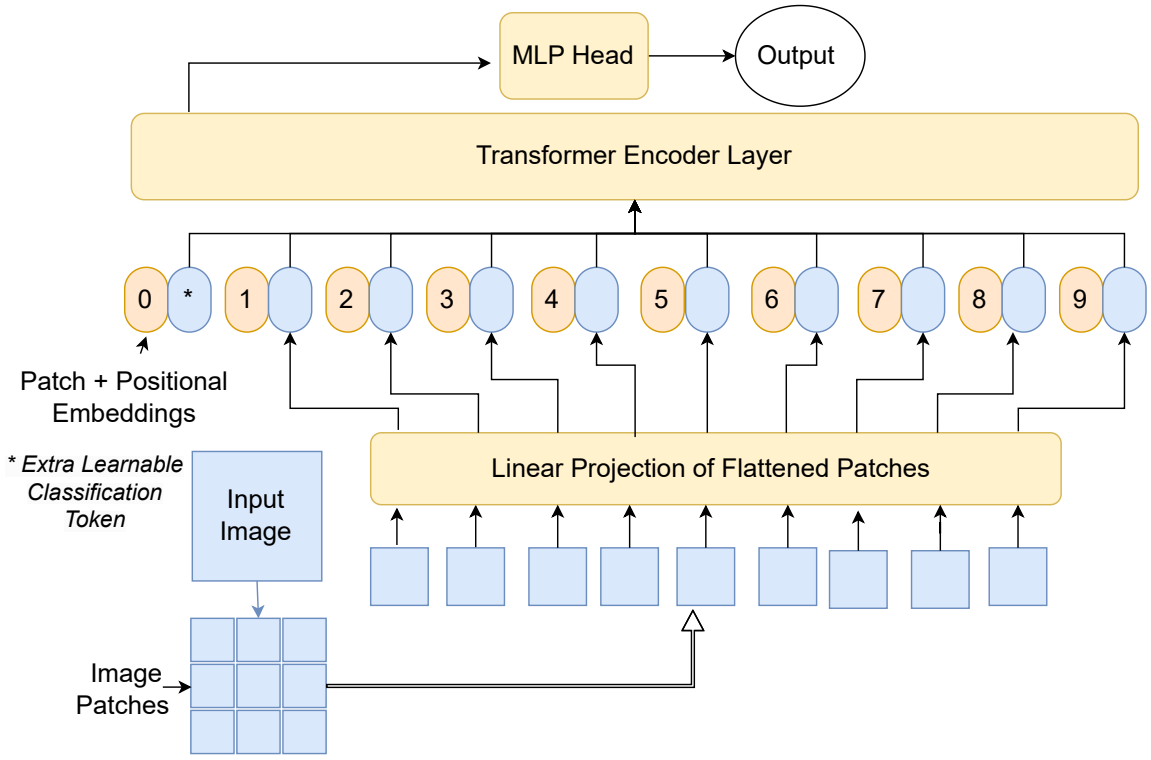


FIGURE 2.5: Scheme of the general architecture of the vision transformer (ViT). The input image is split into small patches and flattened into a sequence of 1D vectors and combined with positional encoding. The numbered circular patches represent the position encoding, and the counterpart represents the flattened 1-D sequence of the image patches. The combined 1-D sequence is passed to the transformer layers. The extra learnable class embedding has encoded the class of the input image after being updated by self-attention and passed it on to an MLP head to predict the output. Image taken from Thuruthipilly et al. (2024b)

(2020) and Hannon et al. (2023) have used DL models pre-trained on the ImageNet data set for the classification of star clusters.

One fact that should be considered when applying transfer learning from one survey to another is the difference in the instruments used in these surveys. For example, two surveys with different zero points will have different pixel values for the same astronomical object in counts/second units. Similarly, various instruments will have different point spread functions (PSF) and pixel scales, which will affect the pixel values of the same source observed by each instrument. Hence, a model trained on one survey could not be directly used on another survey and needs some standardisation. One of the common trends is to re-scale the input image within a specific range (i.e. 0 to 1 or -1 to 1). This also ensures that the input values are not very small, which would compromise the performance of the DL model (de Amorim et al. 2023). Another obstacle to applying transfer learning from surveys to surveys is the different types of instrumental artefacts present in each survey.

For a detailed discussion on transfer learning and its applications, I encourage the reader to see Vilalta (2018) or Hosna et al. (2022).

2.7 Metrics for evaluating deep learning models

The DETR and ViT models described in this chapter are extensively used in my work for the identification of SGLs and LSBGs. This is discussed in detail in Chapter 3 and Chapter 4, respectively. Since I will create various deep learning models in the upcoming chapters, we need some performance metrics to compare the performance of the models created. Here, I give a brief overview of the metrics used generally to compare the performance of the deep learning models.

One commonly used metric to compare different classification models is the accuracy of the classification. For a given model, the classification accuracy is calculated as

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}, \tag{2.9}$$

where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives (also known as type 1 error), and FN is the number of false negatives (also known as type 2 error). Generally, in the literature along with accuracy, the confusion matrix is also used to represent the TP, TN, FP and FN as a matrix to understand how well the model performs in each class. The classification accuracy is a good metric to represent the performance of the model in the cases of balanced datasets. However, for unbalanced datasets, accuracy will not be an effective metric to compare the models. For instance, consider if we had 90 images of galaxies and 10 images of stars. A model predicting all the inputs as galaxies will have a 90% accuracy when tested on this dataset but will not be a good star-galaxy classifier. Hence, we need more metrics such as precision and recall to evaluate the performance of the models which will be described below.

The precision of a model is a useful metric when dealing with unbalanced datasets, as it gives an estimate of how many of the predictions of a model are actually correct. Precision is defined as the ratio between the TP to the total number of predicted positives (TP+FP) and mathematically represented as:

$$\text{Precision} = \frac{TP}{TP + FP}. \tag{2.10}$$

A high value of precision for a label means that the contamination in the predicted label is low. Similar to precision one other useful metric is recall which gives us an estimate of how much of the actual positive labels are correctly identified by the model. In other words, this means that the model makes few false negative errors. The recall is also known as TPR and is defined as the ratio of true predictions to the total number of true cases. Mathematically, recall or TRP can be represented as:

$$TPR = \frac{TP}{TP + FN}. \tag{2.11}$$

The TPR measures how well the classifier detects the true cases from the whole test population of objects. For an ideal classification model, we will have a precision and recall value equal to 1. In practice, models with precision and recall values close to 1 are considered as good classification models.

One other useful metric in comparing unbalanced datasets is known as the false positive rate (FPR) which gives the probability of the models raising a false alarm. The FPR represents the fraction of instances from the negative class incorrectly identified as the

positive class by the model. Mathematically, the FPR can be estimated as:

$$FPR = \frac{FP}{FP + TN}. \tag{2.12}$$

A high value of FPR means that the model is not performing well and predicts too many false positives.

All the metrics defined above depend on a selection threshold to classify an input into a specific class. For instance, in a binary classification model that separates a sample into two classes (class 0 and class 1), a threshold must be chosen to determine the classification of an input predicted with a probability of 0.6. Hence, all the metrics reported above will depend on the chosen threshold. In this scenario, one metric that can be used to estimate the performance of a model is the area under the receiver operating characteristic (AUROC) curve. The receiver operating characteristic curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) as a function of the threshold. The AUROC assesses the overall ability of a classifier to distinguish between classes. A perfect classifier will have AUROC = 1.0 with TPR = 1.0 and FPR = 0.0 for any threshold, whereas a random classifier will have AUROC $\sim$ 0.5 with TPR = FPR for any threshold.

In the upcoming chapters, I will primarily use accuracy, AUROC TPR, and FPR to compare the DL models I have developed for identifying strong gravitational lenses and low-surface brightness galaxies from datasets ranging from simulated data to observations from sky surveys.

3

Finding strong gravitational lenses through self-attention; Study based on the Bologna Lens Challenge

This chapter originally appeared as 'Finding strong gravitational lenses through self-attention; Study based on the Bologna Lens Challenge' by Thuruthipilly, H. et al. 2022, Astronomy & Astrophysics, (Thuruthipilly et al. 2022).

3.1 Introduction

Strong gravitational lensing is the phenomenon by which a distant galaxy or quasar produces multiple highly distorted images because of the gravitational field of the foreground galaxy or a nearby massive astronomical body. Finding and analysing these strong lenses (SLs) is important; they have diverse applications in cosmology and astrophysics, ranging from estimating the Universe's dark matter distribution to constraining the cosmological models (Koopmans et al. 2006; Covone et al. 2009; Collett & Auger 2014; Cao et al. 2015; Bonvin et al. 2016; Bonvin et al. 2017). Consequently, the current and upcoming surveys have given significant attention to detecting strong gravitational lensing systems. For a detailed review of the applications of strong lensing, we refer to Blandford & Narayan (1992) and Treu (2010).

However, for all these analyses a large sample of SLs is required. Unfortunately, only a few hundred lensing systems have been detected and confirmed by the present astronomical surveys to date. One of the largest lens catalogues available now is from the Sloan Lens ACS Survey (SLACS), with only 130 observed lenses (Bolton et al. 2008). With the upcoming era of advanced missions such as Euclid (Euclid Collaboration et al. 2022b) and LSST (Ivezić et al. 2019; Verma et al. 2019), the number of observable SLs is expected to reach 10^5, which should be identified from around 10^9 astronomical objects. Similarly, the number of new SLs expected to be in the Square Kilometre Array (SKA) survey will have similar orders of magnitude (McKean et al. 2015). To analyse the enormous amount of data produced from the present and future large-scale surveys, various methods have been tried out, including crowd science (Marshall et al. 2016) and semi-automated methods, for example, arc detectors (Lenzen et al. 2004; Cabanac et al. 2007). However, these methods have only had minor success and were too time-consuming to be a practical proposition. Hence, the situation demands better and more effective alternative approaches to detect SLs in future large-scale surveys.

It is worth mentioning that the advancements in artificial intelligence (AI) have opened up a plethora of opportunities and have been widely applied in astronomy and astrophysics (e.g. galaxy classification by Pérez-Carrasco et al. 2019, supernova classification by Cabrera-Vives et al. 2017, and lens modelling by Pearson et al. 2019). A particular class of deep-learning models known as convolutional neural networks (CNNs) has recently been shown to work exceptionally well in finding SLs. Hence, developing deep-learning-based algorithms to detect SLs from large-scale surveys is an actively investigated area now (Lanusse et al. 2017; Schaefer et al. 2018; Davies et al. 2019; Chianese et al. 2020). For instance, Jacobs et al. (2017) applied CNNs to the data from the Canada-France-Hawaii Telescope Legacy Survey (CFHTLS) to find SLs, and numerous other successful attempts of finding potential SL candidates from the Kilo Degree Survey (KiDS) have been reported (Petrillo et al. 2017, 2019; Petrillo et al. 2019; He et al. 2020; Li et al. 2020). Likewise, various groups have successfully used CNNs to identify strong lens galaxy-scale systems from large-scale surveys, such as the Dark Energy Survey (DES) (Jacobs et al. 2019; Rojas et al. 2022), the Dark Energy Spectroscopic Instrument Legacy Imaging Surveys (Huang et al. 2020, 2021), and Pan-STARRS (Cañameras et al. 2020), and from comparatively small-scale surveys, such as VOICE (Gentile et al. 2021).

An exciting feature of the CNNs is that they can directly take the image as the input and learn the image features, making them one of the most popular and robust architectures currently being used. Generally, the learning capacity of a neural network increases with the number of layers in the network. The network can then learn the low-level features with the first layers and then learn more complex features with increasing depth (Russakovsky et al. 2015; Simonyan & Zisserman 2015). However, increasing the layers in the neural network will result in higher complexity, which in turn may lead to overfitting (Hawkins 2004). In addition, the gradient of the cost function decreases exponentially. Eventually, it vanishes for very deep networks, commonly called the vanishing gradient effect (Hochreiter 1991; Hochreiter et al. 2001). These two problems meant that creating very deep Convolution Networks was a challenging task.

However, the recently introduced idea of residual learning tackles these problems by introducing skip connections between the input and output of a few convolution layers (He et al. 2016). As a result, the CNN learns the difference between the inputs and outputs rather than their direct mapping. Due to the skip connections, the gradients can reach deeper layers, thus tackling the vanishing gradient effect. Recently, He et al. (2015) were able to build models as deep as 1000 layers while increasing classification accuracy for the ImageNet Large-Scale Visual Recognition Challenge 2015. However, the scientific community are constantly looking for alternative and simple solutions that can outperform the existing models with reasonable computational cost.

Recently, there was a breakthrough in natural language processing (NLP) through the introduction of a new self-attention-based architecture known as the transformers (Vaswani et al. 2017). Since then, there have been attempts to adapt the idea of self-attention to build better image processing models (Parmar et al. 2019; Zhao et al. 2020; Tan et al. 2021). The basic idea behind the transformer architecture is the attention mechanism, which has also found a wide variety of applications in machine learning (Zhang et al. 2019; Fu et al. 2019). In the case of NLP, self-attention correlates with different positions of a single sequence in order to calculate a representation of the sequence. Recently, Facebook Inc. (Carion et al. 2020) and Google Brain (Dosovitskiy et al. 2021) have been able to surpass the existing image recognition models with transformer-based architectures. To our best knowledge, transformer-based models have not been employed in astrophysics yet. In this paper, we explore the possibilities of this new architecture for detecting strongly gravitationally lensed systems.

We implemented various transformer models similar to the architecture described in

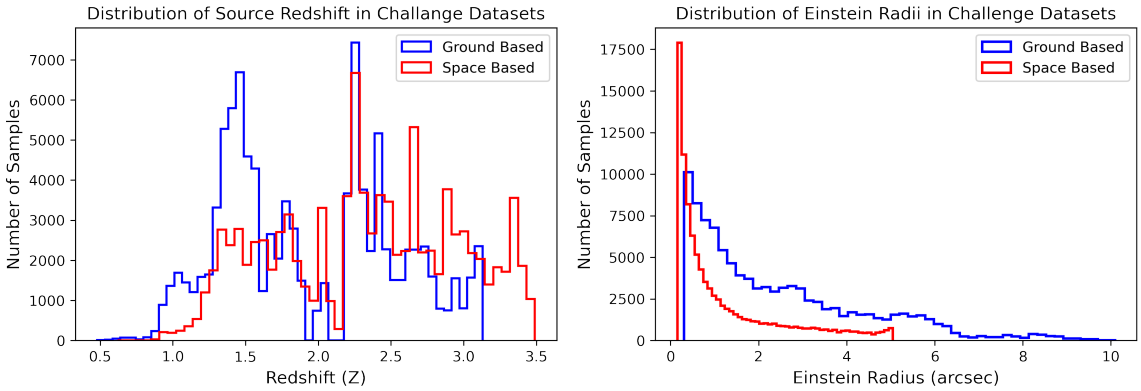


FIGURE 3.1: Distributions of source redshifts and Einstein radii (in arcsec) of simulated gravitational lenses in the Bologna Lens Challenge.

2.5.2 to find the gravitational lenses from the Bologna Lens Challenge. We also compared the performance of the transformer models with created CNNs, and the CNNs participated in the challenge. The main objective of our study was to explore how suitable the transformers are for finding strong lenses and how to optimise the performance of transformers. From our analysis, we found that the transformer models perform better than the CNN models compared. We were able to beat the top TPR_0 and TPR_{10} score (two metrics of evaluation for the Bologna Challenge) by a significant margin and reach the top AUROC reported during the challenge.

The paper is organised as follows. Section 2 briefly describes the data we used to train our models. Section 3 provides a brief overview of the methodology used in our study, including information on how the models were trained. The results of our analysis are presented in section 4. A detailed discussion of our results with a brief review of the performance of transformers compared to the CNN models that participated in the challenge is presented in section 5. Section 6 concludes our analysis by highlighting the advantages of the encoder models over CNN models.

3.2 Data

The data used in this study is from the Bologna Strong Gravitational Lens Finding Challenge (Metcalf et al. 2019). The challenge consisted of two different challenges that could be registered independently. The first challenge was designed to mimic the datasets from surveys such as Euclid consisting of single-band images. The second challenge was designed to resemble data from ground-based detectors with multiple bands. It was roughly modelled on the data from the Kilo-Degree Survey (KiDS) reported in de Jong et al. (2013). However, the simulated images did not strictly mimic the surveys; they were only employed as references to set noise levels, pixel sizes, sensitivities, and other parameters. The distributions of source redshift and Einstein radii in the challenge datasets are shown in Fig. 3.1. The challenge was opened on November 25, 2016, and closed on February 5, 2017. Surprisingly, automated methods such as CNN and SVM showed far better results than human inspection. During the challenge, these methods were able to classify the images with high confidence where a human would have doubt.

The mock images for the challenge were created using Millennium simulation and GLAMER lensing code (Boylan-Kolchin et al. 2009; Metcalf & Petkova 2014). Sources from the Hubble Ultra Deep Field (UDF) decomposed into shapelet functions were used to create the lensed background objects. There were 9,350 such sources with redshifts and

separate shapelet coefficients in four bands. The visible galaxies associated with the lens were simulated using an analytic model for the surface brightness of these galaxies. In particular, the Sérsic profile: $I(R) = I_0 \exp -kR^{1/n_s}$ was used. The parameters employed to simulate the galaxies were the total magnitude, the bulge-to-disc ratio, the disc scale height, and the bulge effective radius. The magnitude and bulge-to-disc ratio are a function of the passband. Each galaxy was provided with an inclination angle between 0° and 80° and random orientation. An elliptical Sérsic profile describes the bulge with an axis ratio randomly sampled between 0.5 and 1. The Sérsic index, n_s, is given by

$$log(n_s) = 0.4log\Big[max\Big(\frac{B}{T}, 0.03\Big)\Big] + 0.1x, \tag{3.1}$$

where x is a uniform random number between -1 and 1 and B/T is the bulge-to-total flux ratio. The median redshift of sources in the space-based catalogue was $z_s = 2.35$ and in the ground-based catalogue it was $z_s = 1.81$.

3.2.1 Space-based

The images for the space-based detector were simulated by Metcalf et al. (2019) to roughly mimic the observations by the Euclid telescope in the visible channel. Metcalf et al. (2019) set the pixel size as 0.1 arcsec, and applied a Gaussian point spread function (PSF) with an FWHM of 0.18 arcsec to simulate the images. The reference band for background and foreground galaxies was the SDSS i, overlapping with the broader Euclid VIS band. The training set consisted of 20,000 images, and the challenge set consisted of 100,000 potential lens candidates.

3.2.2 Ground-based

The ground-based images consisted of simulated images from four bands $(u, g, r,$ and $i)$, and the reference band was the r band. In the challenge set, 85% of the images were purely simulated. The other 15% were actual images chosen from a preliminary sample of bright galaxies directly from the KiDS survey. These real images were added to the challenge set for more realism. Some images had masked regions where removed stars, cosmic rays, and bad pixels were present. The noise for the mock images was simulated by adding normally distributed numbers with the variance given by the weight maps from the KiDS survey. The example images of a mock simulated lens for the challenge are shown in Fig. 3.2. For a detailed review of how the data was created, we refer to Metcalf et al. (2019).

An exciting result reported from the challenge was that colour information was crucial for finding strong lenses. All the methods that participated in the challenge performed better on the data from the ground-based observatories, which had four photometric bands (u, g, r, i), than on the data from the space-based detectors, which had a single band. Consequently, it was advised by Metcalf et al. (2019) to add even low-resolution information from other instruments or telescopes to the higher resolution data in one band to improve the detection rates significantly. In other words, multiple bands make a significant difference, and future surveys will perform better if they have information provided in multiple bands. Hence, for our study, we chose the data from the ground-based observatories with four photometric bands (u, g, r, i) to study the attention-based models' ability to detect strong gravitational lenses. Since we are also interested in exploring the transformer architecture's optimisation and analysing the transformers' performance, a better data structure was preferred to compare transformer models.

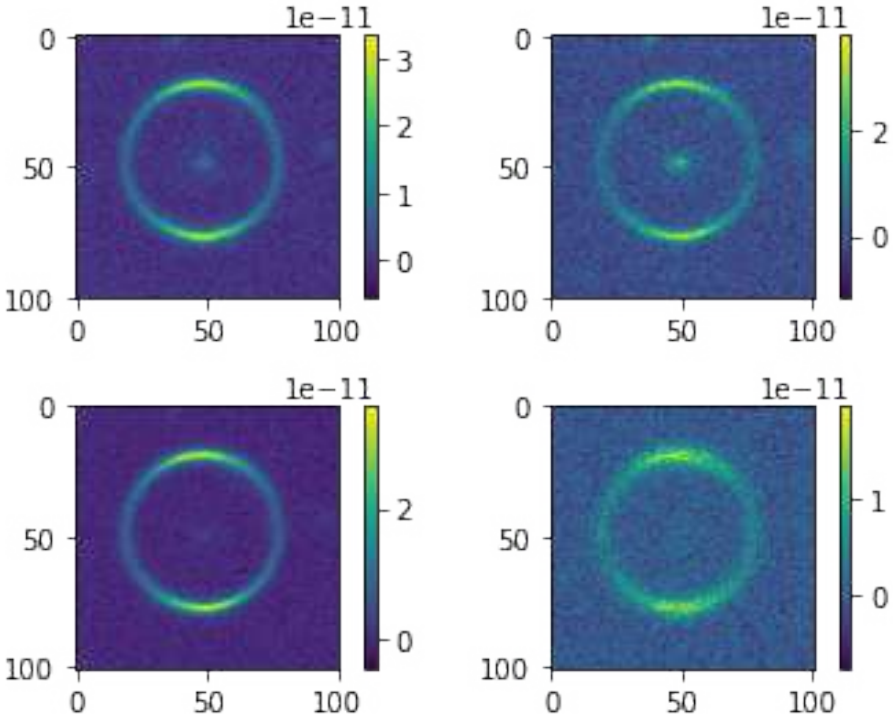


FIGURE 3.2: Typical image of a mock simulated lens for the challenge. Bands are shown in the following order: u (top left), g (top right), r (bottom left), and i (bottom right).

3.2.3 Data pre-processing

The simulated datasets of the Ground-Based Bologna Strong Gravitational Lens Finding Challenge were provided in the FITS format and were available for download to the public [1]. The challenge datasets contained 100,000 potential strong lens candidates, and the training set contained 20000 images along with other information, such as the Einstein area in rad^2 and number of pixels in the lensed image above $1 \times \sigma$. In this work we did not use additional information about the images. We only used whole images (101×101 pixels) in all four photometric bands (u, g, r, i) as an input to the model and information about the lens present or not as the desired output for training the models. During training the 20000 images were split into two parts. We used a dataset of 18,000 to train the network, which was used for validation. Before training the models, each image was re-scaled and rotated by $n\pi/2$, where $n \in (0, 1, 2, 3)$, to enrich the dataset.

3.3 Methodology

I created 21 Transformer models similar to the model described in Sect. 2.5.2 which we call encoder models because of its resemblance to the encoder part of the model presented by Carion et al. (2020). I created different encoder models with different structures to study how the hyperparameters in the encoder will affect the model's performance. We used the exponential linear unit (ELU) function as the activation function for all the layers in these models. We initialise the weights of our model with the Xavier uniform initialiser, and all layers are trained from scratch by the ADAM optimiser with the default exponential decay rates (Glorot & Bengio 2010a; Kingma & Ba 2015).

[1] http://metcalf1.difa.unibo.it/blf-portal/gg_challenge.html

3.3.1 Lens detector

Among the created encoder models, the best performance was given by the encoder model that uses a CNN backbone similar to the LASTRO-EPFL model, from the Bologna Lens Challenge (Schaefer et al. 2018). Here we present the two best architectures: Lens Detector 15 and Lens Detector 16, which outperformed all the other models during our analysis.

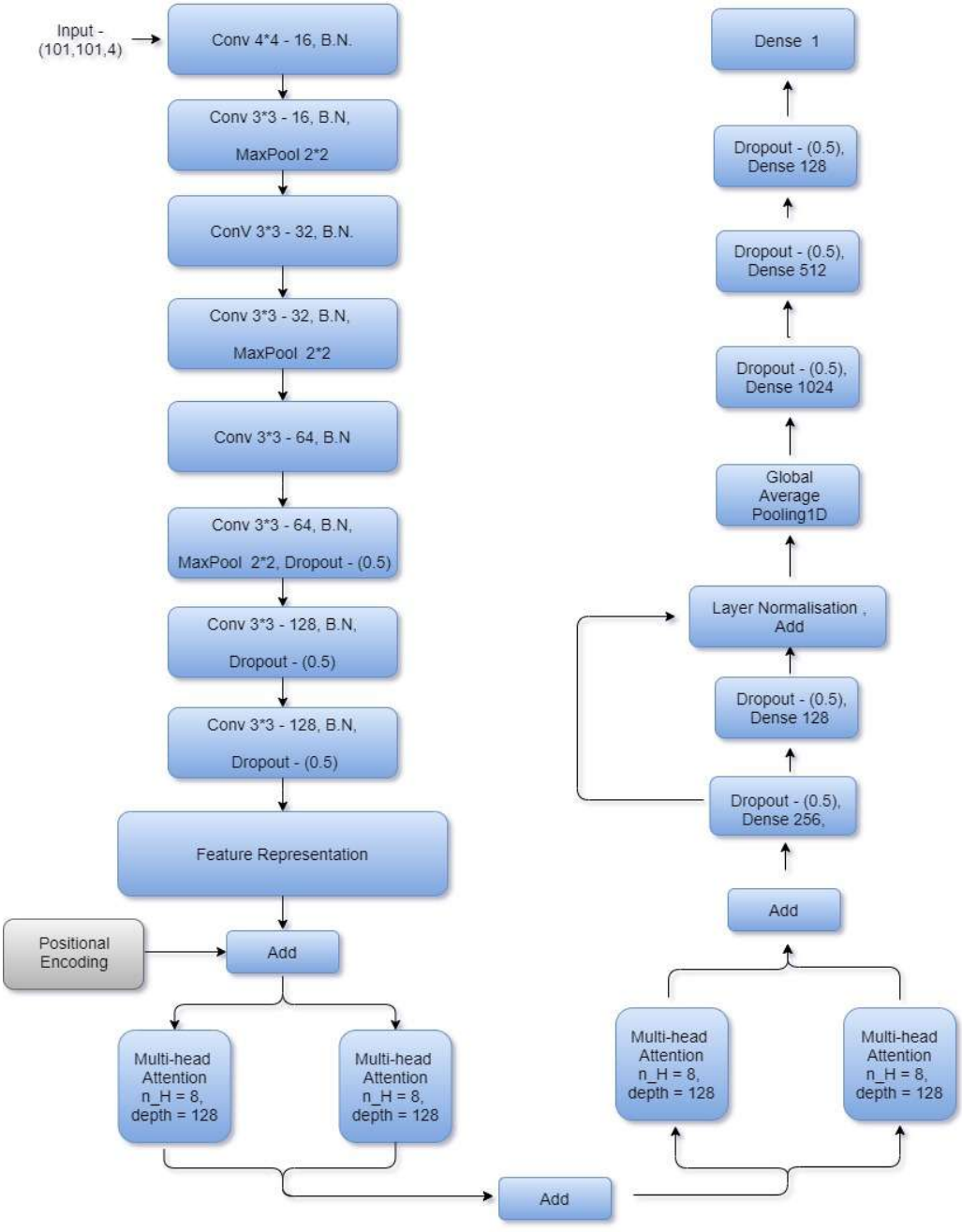


FIGURE 3.3: Scheme of the architecture of Lens Detector 15

Lens Detector 15 was first trained for 300 epochs with an initial learning rate of $\alpha = 10^{-4}$ and again trained for another 100 epochs starting with a learning rate of $\alpha = 10^{-5}$. This version of the lens detector gave high scores in all three evaluation metrics for the

challenge. The architecture of Lens Detector 15 is given in Fig. 3.3. In the spirit of reproducible research, our code for Lens Detector 15 is publicly available[2].

Lens Detector 16 was created by stacking two Lens Detector 15 models in parallel and combining their outputs through an additional dense layer connected to a single neuron to give the output. The architecture of Lens Detector 16 is shown in Fig. 3.4. Lens Detector 16 was first trained for 100 epochs with an initial learning rate of $\alpha = 10^{-4}$ and again trained for another 100 epochs starting with a learning rate of $\alpha = 5 \times 10^{-5}$. Furthermore, the model was trained for 50 epochs with $\alpha = 10^{-5}$ and after that with $\alpha = 5 \times 10^{-6}$ for another 200 epochs.

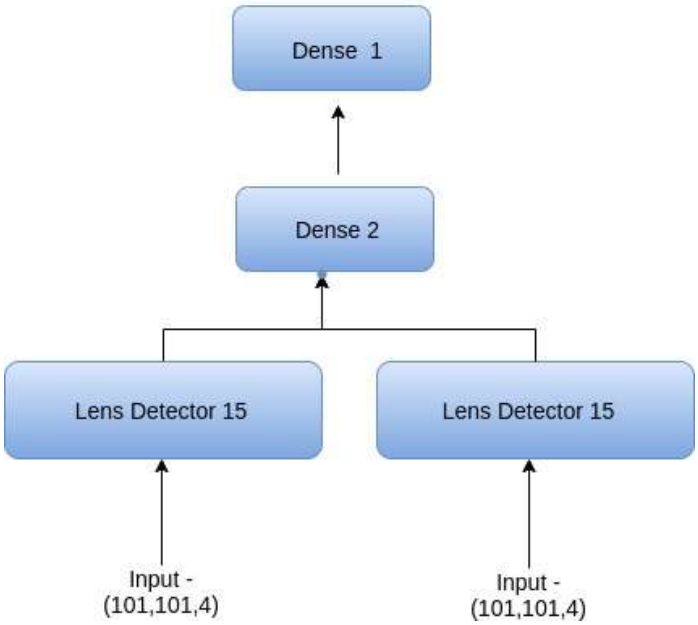


FIGURE 3.4: Scheme of the architecture of Lens Detector 16

We created the Space Lens Detector model to identify strong lenses from the space-based dataset. The Space Lens Detector has a similar structure to Lens Detector 15. The only difference is the use of four heads in the encoder layers. The model was trained with an initial learning rate of $\alpha = 10^{-4}$ using the ten-fold validation and iterating for 20 epochs in each fold.

3.3.2 Metrics for evaluation

For comparing the models created, we use the metrics described in section 2.7. For the Bologna Lens Challenge, the participants were instructed to optimise AUROC rather than the accuracy. In addition, two more figures of merit were also considered for the competition, which are TPR_0 and TPR_{10}. The TPR_0 is defined as the highest TPR reached, as a function of the p threshold, before a single false positive occurs in the test set of 100,000 cases. This is the point where the ROC meets the FPR = 0 axis. If the classifier assigns a high probability for a non-lensed image to be a lens, even for one case, the TPR_0 will go low. This means that the TPR_0 parameter measures the confidence in the purity of the samples identified by a model. Similarly, the TPR_{10} is defined as the TPR at the point where fewer than ten false positives are made, which is also a measure of confidence in the true samples mined out by a model with a slight impurity. A high TPR_0 and TPR_{10} indicate that the classifier can clearly distinguish between lensed and non-lensed images.

[2] https://github.com/hareesht23/Lens-Detector

3.4 Results

We created 5 convolution models to use as the backbones for the encoder models and 21 encoder models to study how the hyperparameters of the encoder layer affect the performance. Since each architecture was implemented as a regression model, a probability of 0.5 was set as the threshold for classifying an image as a lens or not. Thus, input images with a prediction value less than 0.5 were classified as non-lensed images labelled zero and vice versa. Table 3.1 describes the architecture and total accuracy, AUROC, TPR_0, and TPR_{10} of all created models.

Model name	Model structure	Accuracy	AUROC	TPR_0	TPR_{10}
CNN 1	5 CNN Layers	88.21	0.951	0.000	0.07
CNN 2	4 CNN Layers	86.74	0.915	0.000	0.4
CNN 3	8 CNN Layers	88.51	0.968	0.033	0.37
CNN 4	3 CNN Layers	88.49	0.956	0.000	0.68
CNN 5	25 CNN Layers	91.26	0.974	0.004	0.004
Lens Detector 1	CNN 1+1 H_{16}+1(E)	89.57	0.961	0.000	0.643
Lens Detector 2	CNN 2 + 1 H_{16} + 1(E)	88.13	0.950	0.001	0.001
Lens Detector 3	CNN 2 + 2 H_{16} + 1(E)	88.00	0.962	0.018	0.018
Lens Detector 4	CNN 2 + 2 H_{32} + 1(E)	88.12	0.952	0.121	0.124
Lens Detector 5	CNN 2 + 4 H_{64} + 2 (E)	88.46	0.955	0.125	0.133
Lens Detector 6	CNN 2 + 4 H_{128} + 4(E)	89.51	0.957	0.003	0.004
Lens Detector 7	CNN 3 + 8 H_{128} + 2(E)	91.45	0.968	0.000	0.410
Lens Detector 8	CNN 4 + 2 H_{384} + 2 (E)	89.43	0.954	0.000	0.758
Lens Detector 9	3 CNN Layers + 2 H_{384} + 2 (E)	89.61	0.959	0.000	**0.789**
Lens Detector 10	5 CNN Layers + 8 H_{128} + 2 (E)	90.58	0.970	0.180	0.23
Lens Detector 11	5 CNN Layers + 8 H_{128} + 4 (E)	90.45	0.966	0.219	0.34
Lens Detector 12	8 CNN Layers + 8 H_{128} + 4 (E)	89.82	0.960	0.040	0.680
Lens Detector 13	8 CNN Layers + 8 H_{128} + 4 (E)	91.94	0.975	0.175	0.525
Lens Detector 14	8 CNN Layers + 8 H_{128} + 4 (E)	91.95	0.975	0.002	0.539
Lens Detector 15	8 CNN Layers + 8 H_{128} + 4 (E)	**92.99**	0.978	0.140	0.48
Lens Detector 16	16 CNN Layers + 8 H_{128} + 8 (E)	90.97	0.962	**0.225**	0.24
Lens Detector 17	16 CNN Layers + 8 H_{128} + 8 (E)	92.19	0.973	0.00	0.717
Lens Detector 18	16 CNN Layers + 8 H_{128} + 8 (E)	92.09	0.976	0.113	0.590
Lens Detector 19	16 CNN Layers + 16 H_{128} + 8 (E)	90.03	0.961	0.114	0.115
Lens Detector 20	25 CNN Layers + 8 H_{128} + 4 (E)	91.26	0.973	0.212	0.223
Lens Detector 21	8 CNN Layers + 8 H_{128} + 4 (E)	92.79	**0.98**	0.00	0.64

TABLE 3.1: Table comprising the architecture, accuracy, AUROC, TPR_0, and TPR_{10} of all the models in chronological order of creation. The encoder models are named 'Lens Detector' followed by a number. The model structure describes if the model uses transfer learning in the CNN backbone or not. Generally, the term 'XH_Y' in the model structure means there are x heads with dimension y in one encoder layer. Similarly, the term '$Z(E)$' denotes that there are Z encoders in the structure.

Among the created encoder models the highest accuracy was achieved by Lens Detector 15 and the highest AUROC, TPR_0, and TPR_{10} were achieved by Lens Detector 21, Lens Detector 16, and Lens Detector 9, respectively. From the presented models here, we would like to highlight Lens Detector 15 as the best model since it performs well in all categories and has the highest classification accuracy. The receiver operator characteristic (ROC) curve of Lens Detector 15 is shown in Fig. 3.5. Similarly, Lens Detector 13, Lens

Detector 18, and Lens Detector 20 can also be considered highly performing classifiers. All of these models scored an AUROC equivalent to the second-best model that participated in the challenge and a better TPR_0 and TPR_{10} compared to all other models that participated in the challenge. The ROC curves of all the encoder models presented in Table 3.1 are displayed in Fig. 3.6.

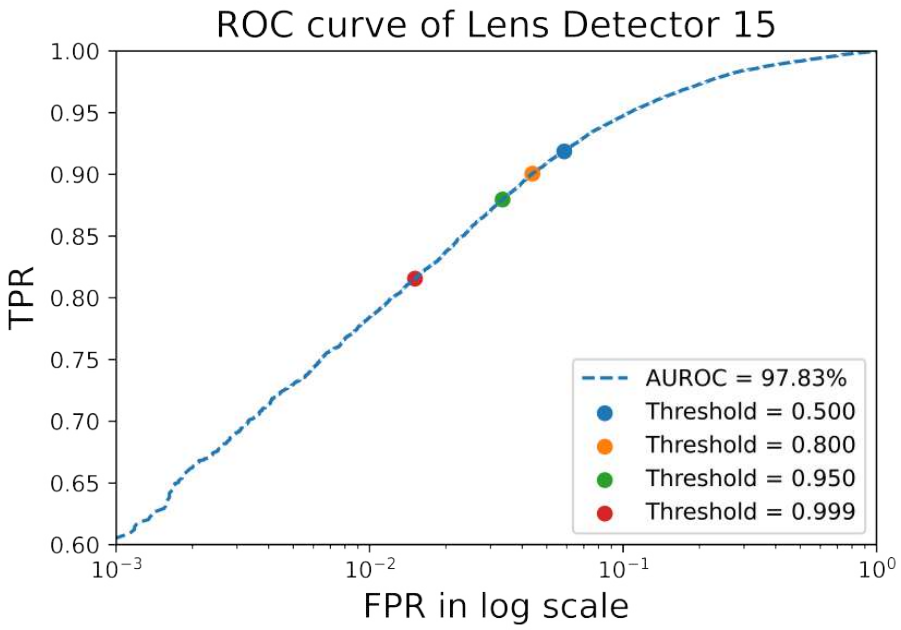


FIGURE 3.5: Receiver operating characteristic (ROC) curve of Lens Detector 15.

Even though we chose 0.5 as the threshold for identifying a candidate as a strong lens, in reality, such a threshold is not practical since the number of lenses to visually inspect after the run of the network could be unrealistically high. In order to validate the performance of Lens Detector 15, we plotted the confusion matrix of the lens detector with varying thresholds (see Fig. 3.7). For the ground-based data, we were able to mine out more than 80% of the true lenses with a threshold as high as 0.999.

To identify strong lenses from the space-based dataset, we created the Space Lens Detector, which scored an AUROC = 0.925. The corresponding AUROC is greater than the second-best AUROC reported in the Bologna Lens Challenge and a little below the top AUROC (0.93) of the Bologna Lens Challenge. The Space Lens detector scored a TPR_0 of 0.039 and TPR_{10} of 0.166, which is comparable to the performance of the other machine learning techniques that participated in the challenge. Since it is clear that the attention-based encoder models can identify the SLs from single-band images, we did not try to improve the scores further. The ROC curve of the Space Lens Detector is shown in Fig. 3.8.

Similarly to the models that participated in the challenge, encoder models also performed better on the data from the ground-based observatories, which had four photometric bands (u, g, r, i), compared to the data from the space-based detectors, which had a single band. Our results support the argument presented in Metcalf et al. (2019) that multiple bands make a significant difference and improve the detection.

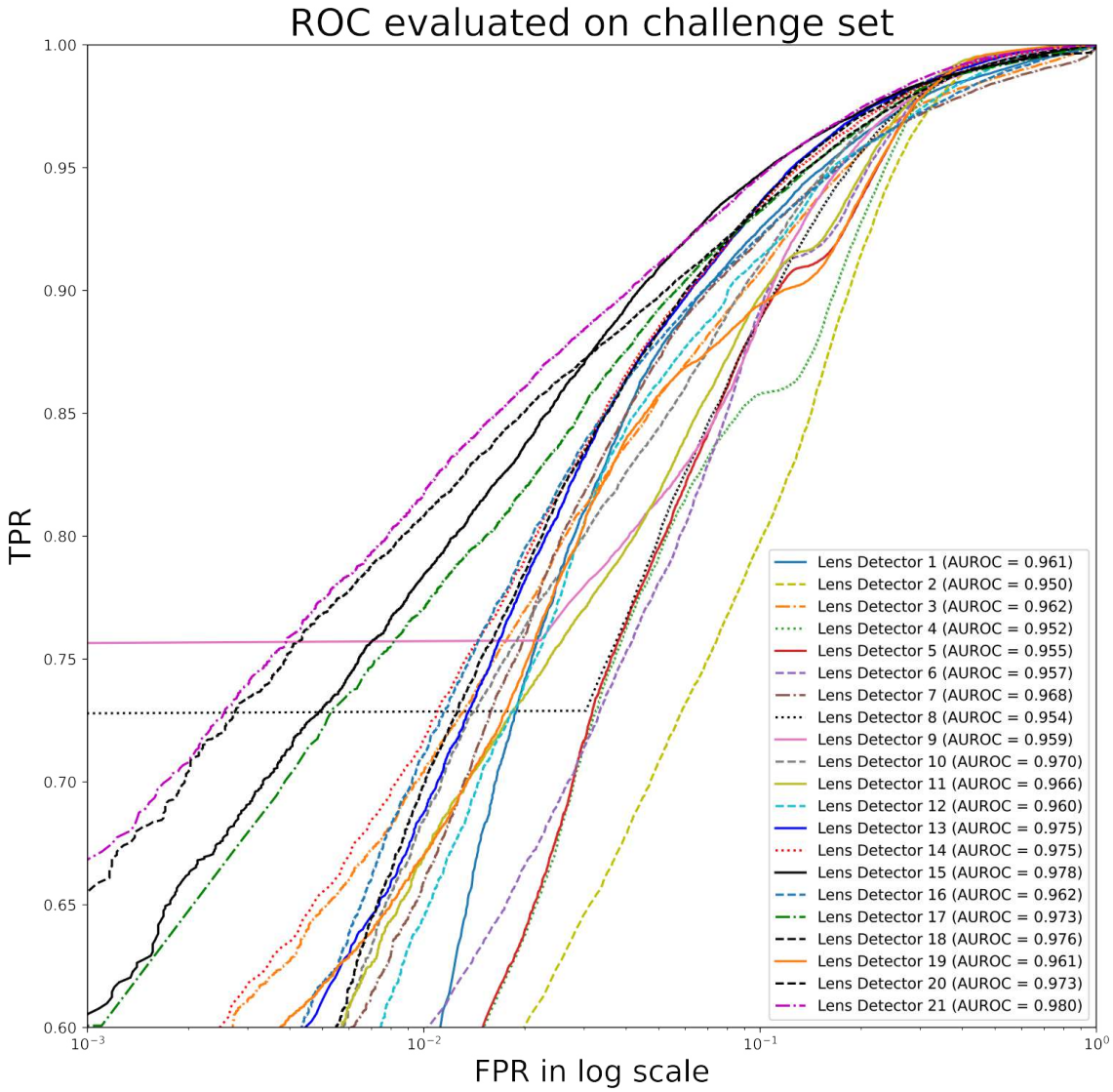


FIGURE 3.6: ROC curves of all the encoder models. The inset reports the AUROC score of each model in order to facilitate comparison of the models.

3.5 Discussion

3.5.1 Transformers and models from the Bologna Lens Challenge

The Bologna Lens Challenge was intended to improve the efficiency and biases of tools used to find strong gravitational lenses on galactic scales. It was clear from the challenge that automated methods such as CNNs and SVM have a clear advantage compared to conventional methods. The performance of all these models was evaluated using AUROC, TPR_0, and TPR_{10} scored on the challenge set. An SVM model named Manchester SVM won the competition in the TPR_0 category with a score of 0.22. (Metcalf et al. 2019; Hartley et al. 2017). The model named CMU Deep Lens received an AUROC of 0.981 and a TPR_{10} score of 0.45, the highest in their respective categories, and thus won the competition (Metcalf et al. 2019; Lanusse et al. 2017). Another variant of the model, named CMU Deep Lens, also received an AUROC of 0.98 during the challenge. These models were 46 layers deep ResNets with around 23 $\times 10^6$ parameters (Lanusse et al. 2017). Another model worth mentioning from the challenge is LASTRO EPFL, an eight-layer CNN that won the

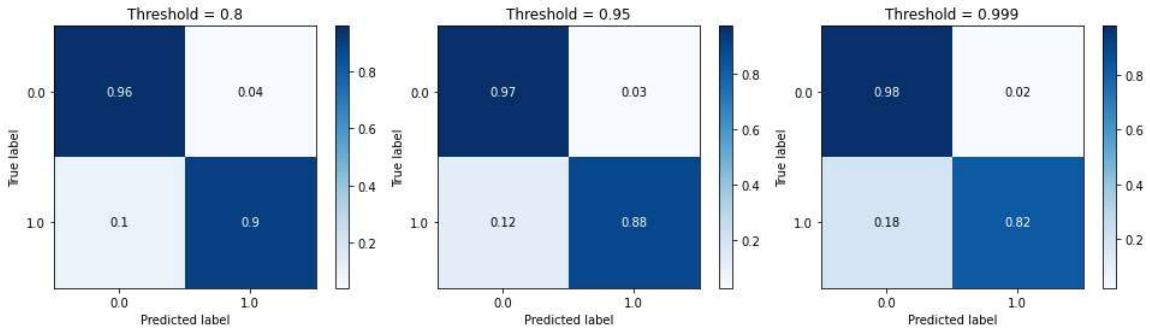


FIGURE 3.7: Confusion matrix of Lens Detector 15 plotted for various thresholds. Class 0 represents the non-lensed images, and Class 1 represents the lensed images. The lower right square in each confusion matrix represents the true positives for which Lens Detector 15 identified strong lenses correctly. The upper left square in each confusion matrix represents the true negatives for which Lens Detector 15 identified non-strong lenses correctly. The lower left square in each confusion matrix represents the false negatives or the missed true lenses by Lens Detector 15. The upper right square in each confusion matrix represents the false positives or the non-lenses identified by Lens Detector 15 as strong lenses.

competition for the space-based dataset in the AUROC category. For a detailed look at the models that participated in the challenge we refer to Hartley et al. (2017) for the SVM, Lanusse et al. (2017) for CMU-DeepLens, and Schaefer et al. (2018) for LASTRO EPFL.

We would like to compare the performance of these models to the performance of the encoder models to exhibit the advantages of encoder models over the CNNs and SVM models. As mentioned earlier, we focused on the data from ground-based observatories. Here we are only comparing the performance of the created encoder models and that of the models that participated in the challenge only for the ground-based observatories data. The values reported in Metcalf et al. (2019) are used here for the comparison.

During the challenge, the TPR_0 was used to strongly penalise the classifiers with discrete ranking because their highest classification level was not conservative enough to eliminate all false positives, and they were likely to get $TPR_0 = 0$. For the other models that participated in the challenge, maximising the TPR_0 was a tough challenge, also for encoder models. However, the encoder models performed very well compared to the CNN models that participated in the challenge. The results of TPR_0 for the top three encoder models and the top three models that participated in the challenge are listed in Table 3.2. We would like to note two models; Lens Detector 16 achieved a TPR_0 of **0.225** and Lens Detector 11 reached a TPR_0 of 0.219, which are very high compared to the CNNs that participated in the challenge.

Name	AUROC	TPR_0	TPR_{10}	Model type
Lens Detector 16	0.962	0.225	0.24	Transformer
Manchester SVM	0.93	0.220	0.35	SVM/Gabor
Lens Detector 11	0.966	0.219	0.34	Transformer
Lens Detector 15	0.978	0.140	0.48	Transformer
CMU-DeepLens				
Resnet-ground3	0.98	0.09	0.45	CNN
LASTRO EPFL	0.97	0.07	0.11	CNN

TABLE 3.2: Comparison of encoder models and models that participated in the Bologna Lens Challenge, listed in decreasing order of TPR_0.

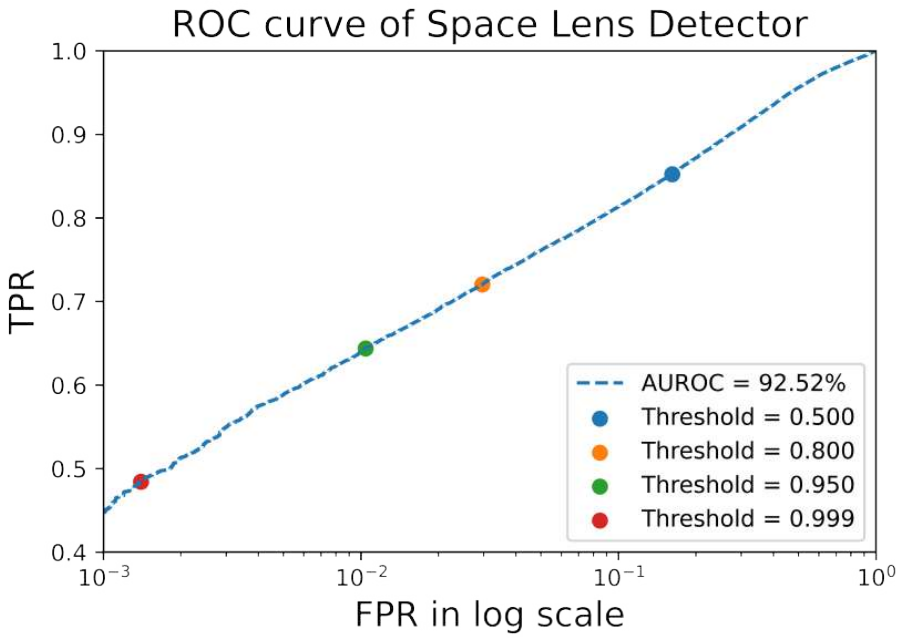


FIGURE 3.8: ROC curve of the Space Lens Detector.

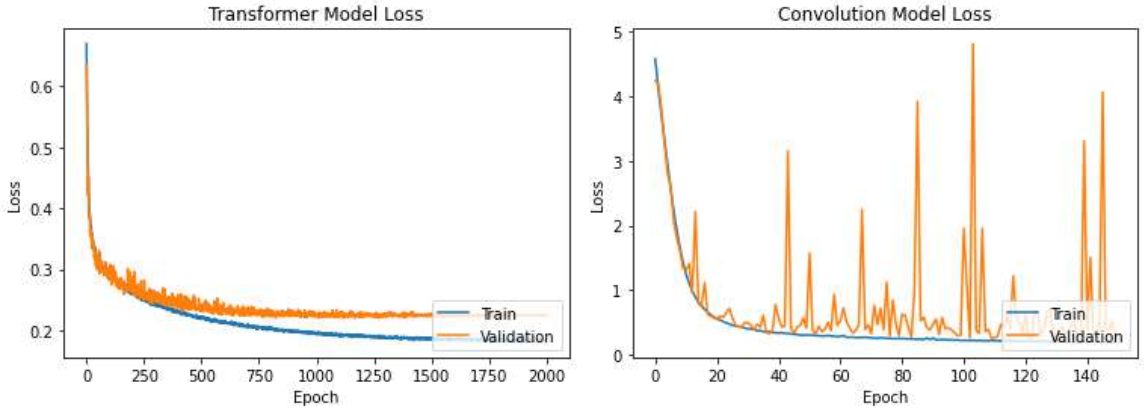


FIGURE 3.9: Variation of loss function with epochs for Lens Detector 13 and CNN 3, respectively. Lens Detector 13 uses CNN 3 as its CNN backbone.

The next parameter used to evaluate the models in the challenge was TPR_{10} for which the encoder models showed a high range of supremacy over all other models that participated in the challenge. Particularly, Lens Detector 9 achieved $TPR_{10} = 0.79$. The results of TPR_{10} for the top three encoder models and the top three models that participated in the challenge are listed in Table 3.3. Three of our models were able to score a TPR_{10} above 0.70, which is very high compared to the top TPR_{10} reported during the challenge. Table 3.1 clearly shows that most encoder models achieved a higher score in this category compared to the other models that participated in the challenge.

Name	AUROC	TPR_0	TPR_{10}	Model type
Lens Detector 9	0.959	0.00	0.789	Transformer
Lens Detector 8	0.954	0.00	0.758	Transformer
Lens Detector 17	0.973	0.00	0.717	Transformer
CMU-DeepLens Resnet-ground3	0.98	0.09	0.45	CNN
Manchester SVM	0.93	0.220	0.35	SVM/Gabor
LASTRO EPFL	0.97	0.07	0.11	CNN

TABLE 3.3: Comparison of encoder models and models that participated in the Bologna Lens Challenge, listed in decreasing order of TPR_{10}.

Now looking at the third parameter of merit used in the Bologna Lens Challenge, which is the AUROC, we can see that Lens Detector 21 was able to reach the highest reported AUROC in the Bologna Lens Challenge (Metcalf et al. 2019). The top three encoder models and the top three models that participated in the challenge that scored the highest AUROC are listed in Table 3.4. However, the CMU-DeepLens, was a 46 layer deep ResNet with around 23 $\times 10^6$ parameters (Lanusse et al. 2017), whereas Lens Detector 21 had only 3 $\times 10^6$ parameters and achieved an AUROC of 0.9809, which is very close to the performance of CMU Deep Lens (AUROC = 0.9814).

Name	AUROC	TPR_0	TPR_{10}	Model type
CMU-DeepLens Resnet-ground3	0.981	0.09	0.45	CNN
Lens Detector 21	0.981	0.00	0.64	Transformer
CMU-DeepLens Resnet-Voting	0.980	0.02	0.10	CNN
Lens Detector 15	0.978	0.140	0.48	Transformer
Lens Detector 18	0.976	0.113	0.59	Transformer
LASTRO EPFL	0.97	0.07	0.11	CNN

TABLE 3.4: Comparison of encoder models and models that participated in the Bologna Lens Challenge, listed in decreasing order of AUROC.

3.5.2 Insights into transformers

An initial glance at the results in Table 3.1 shows that encoder models perform better than CNN models. However, the encoder models depend on the CNN backbone to extract the features, and as a result the performance of the encoder models depends upon the CNN backbone. A detailed look at the results indicates that the encoder model is only good as its CNN backbone. However, the encoder model always performs better than its CNN backbone. For example, the lowest accuracy achieved among the encoder models was for Lens Detector 3, which has better performance than CNN 2. A similar trend can be observed for other encoder models, which use trained CNNs as their backbone and perform better than the trained CNNs. These observations show that the encoder models can achieve accuracy that is better by a small percentage than a CNN with the same number of convolution layers.

However, an interesting question that should be addressed is what happens if we use deeper CNN backbones. We compare the performance of a deeper CNN and an encoder model with a deeper CNN backbone. Model CNN 5 with 25 convolution layers and Lens Detector 20 with the same number of convolution layers give same AUROC (0.97) and the

same accuracy (0.91). We can see that the two models perform equally well. Using self-attention in deeper CNNs did not significantly improve the AUROC or accuracy. However, the TPR_0 score and TPR_{10} score of Lens Detector 20 (0.212 and 0.223, respectively) is higher compared to CNN 5 (0.004 and 0.004, respectively). One probable reason why self-attention does not improve the accuracy and AUROC of deeper CNNs is that the CNN backbone will learn more about the image's micro-scale features in deeper layers. Hence, the model will miss the long-range correlations of the original image found by the encoder layer.

In this section, we speak of the number of hyper-parameters in the encoder layer. On analysing the results from the encoder models, we can see that increasing the number of heads and the depth of the encoder increases the model's performance. However, increasing the number of heads and the depth of the encoder also increases the number of trainable parameters in the model. During the training period, it was found that increasing the number of parameters in the encoder layer helps the model to learn faster. This points to an exciting aspect of the encoder models. The encoder's performance is proportional to the number of trainable parameters in the encoder layer or specifically in the multi-head attention layer. The higher the number of trainable parameters, the better the learning curve and performance. However, the performance saturates beyond a limit for a given CNN backbone.

Another interesting observation was the difference in the number of trainable parameters in the CNNs and self-attention-based encoder models. For example, an eight-layer CNN will have 4×10^6 parameters, whereas the encoder with eight CNN layers and four encoder layers with eight heads each will have 3×10^6 parameters. In CNNs, most of the parameters come from the connections between the flattened output of the last convolution layer to the following dense layer. The weights in these layers help the CNNs to learn the features of the image. However, for a transformer network most of the trainable parameters are in the attention layers, which are only trying to learn the long-range correlations in the data and effectively act as a filter. This is one of the reasons why transformer networks can prevent overfitting. However, transformer models did not show any advantages over CNNs for the time taken to test and train the models.

We also tried different learning rates from different ranges to find the optimal learning rate. We found that increasing the learning rate above 0.001 considerably reduced the performance of the lens detectors. With an initial learning rate of 0.01, the models were not able to learn the features of the lenses from the training set. The optimal learning rate for our model was found to be between $\alpha = 5 \times 10^{-5}$ and $\alpha = 2 \times 10^{-4}$; we chose $\alpha = 10^{-4}$ as the initial learning rate for our models.

Another striking feature worth pointing out is that, unlike convolution layers increasing the number of parameters in the encoder layers has a very slight effect on overfitting the model. Since the self-attention layers act as the filters for features extracted by CNN, an increase in the number of parameters in the encoder layers helps the models to filter the features faster and effectively without causing the overfitting of the model. The effect of self-attention layers in filtering and smoothing the learning curve can also be seen in comparing the loss curve of the CNN and encoder models presented in Fig. 3.9.

We also tried transfer learning by using an already trained CNN as the backbone of the encoder model. Surprisingly, however, the encoder models that do not use transfer learning performed slightly better than the models that use transfer learning. Since a trained CNN model has already learned to extract specific features of an image, the encoder model with that CNN backbone is restricted to minimise the loss function in only a small part of the hyperspace. So, the self-attention layers can only filter the features and improve the accuracy by a small percentage (e.g. CNN 2 86.74% and Lens Detector 2, with CNN 2 as

the backbone, 88.13%). Nevertheless, for a model without transfer learning, there is a possibility for the CNN part in the encoder model to learn more features than a solo CNN about the image and improve the accuracy. For example, Lens Detector 7, which used the trained CNN 3 as its backbone, scored an accuracy of 91.45%. In contrast, Lens Detector 15, which used a CNN backbone similar to CNN 3 and without any transfer learning, scored an accuracy of 92.99%. However, this result cannot be generalised since it also depends on the trained CNN backbone.

An updated version of the CNN models that participated in the challenge and had better scores in every category compared to their previous versions has recently been reported by Magro et al. (2021). They used the same CNNs that participated in the challenge and retrained the networks with different epochs. Even though the models had improved scores, it was evident from the report that the performance of the models is highly dependent on the number of epochs. In other words, the CNNs reported in the Bologna Lens Challenge had lower stability. We have to monitor the training to achieve better results carefully. In contrast, the encoder models are highly stable compared to the CNNs. As shown in Fig. 3.9, we were able to train the encoder models up to 2000 epochs without any sign of overfitting. Interestingly, the fluctuations in the validation loss were very stable up to the end.

3.5.3 Lens detectors for strong lens detection

It is worth pointing out that encoder models can identify SLs and non-SLs better than their CNN counterparts. The probability distribution of finding a lens in the challenge dataset is depicted in Fig. 3.10. The encoder model can assign a probability for an input to be a lens ($p \approx 1$) or non-lens ($p \approx 0$) with greater confidence than the CNN. Furthermore, from Fig. 3.10, it is clear that the transformer models can approximately mimic a perfect classifier by assigning a probability of 0 to non-lensed images and a probability of 1 to lensed images. This feature of the encoder model will be beneficial and applicable in the upcoming large-scale surveys to narrow down the potential lensing systems with great confidence. Figure 3.7 shows that Lens Detector 15 was able to identify 82% of the true lenses with a probability greater than 0.999, which explains the peak near the point 1.0 on the x-axis of Fig. 3.10.

Here, we used all four bands available for training the models. However, the u-band images are usually not used to search for strong lenses because, in the real scenario, the u-band images often have lower image quality. It is also possible that they are not available for fainter galaxies. In the literature, for detecting the strong lenses, the images from the g, r, and i bands are used for training machine learning models. Sometimes along with a three-channel CNN, another single-band CNN is also created, and the combined predictions of these two CNNs are used to shortlist the real lens candidates (Petrillo et al. 2019; Petrillo et al. 2019; Li et al. 2020). To test the adaptability of the encoder model for three bands, we removed the u band and retrained the Lens Detector 15 model from scratch with the images from the g, r, and i bands and named it the 3-band Lens Detector. The retrained Lens Detector 15 achieved an AUROC = 0.974, which is comparable to the AUROC when the u band was present. The ROC curve and the confusion matrix for various thresholds are presented in Fig. 3.11 and Fig. 3.12. Comparing the confusion matrices in Fig. 3.7 and Fig. 3.12, we can see that removing the u band slightly reduces the number of true positives for a given threshold since the information available in the u band was gone. For example, for a threshold = 0.8, Lens Detector 15 identified 90% of the true positives and the 3-band Lens Detector identified 88% of the true positives. This result also validates the argument presented in Metcalf et al. (2019) to include even low-resolution information in one band to improve the detection rates.

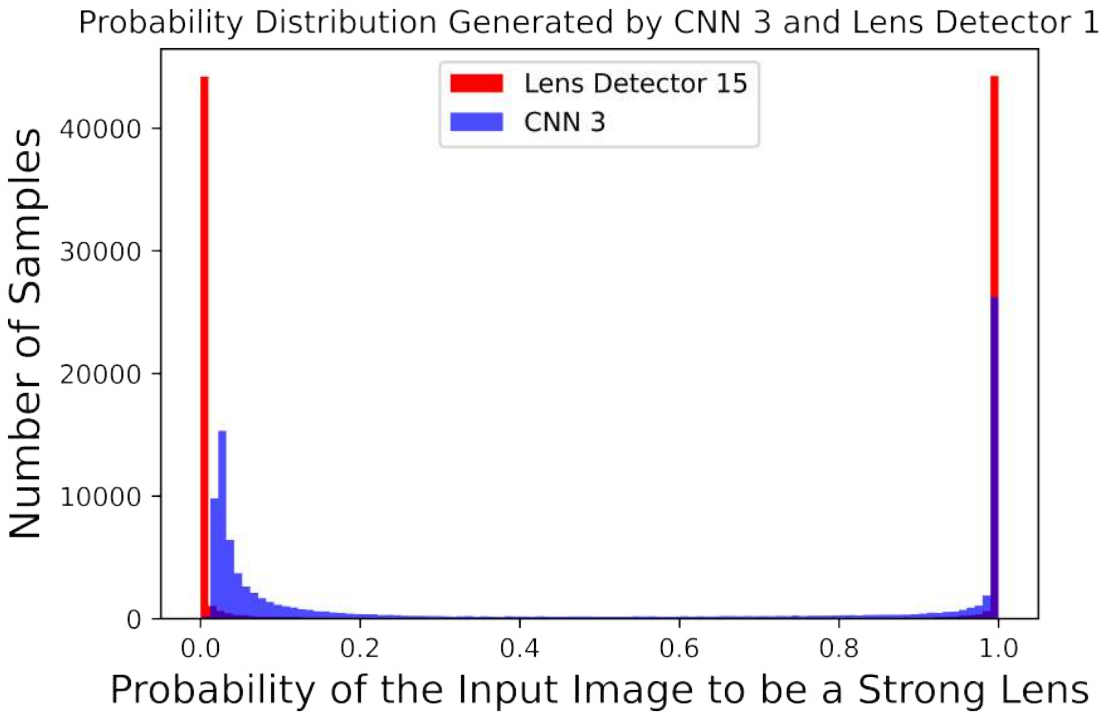


FIGURE 3.10: Comparison of the output probabilities of CNN 3 and Lens Detector 15 for the ground-based challenge dataset. In this histogram values leaning towards zero represent the lack of a strong lens in the image, and values leaning towards one indicate the presence of a strong lens in the image.

Since we already tested the encoder model on the space-based dataset and obtained comparable results with models that participated in the Bologna Lens Challenge, it is clear that encoder models can be adapted for a single-band analysis. However, another interesting question to be investigated is the ability of the encoder models to find strong lenses from a different data distribution than the model has been trained on. We retrained the Space Lens Detector using the data in the *r* band from the ground-based data and tested it on the space-based challenge dataset to investigate this aspect. The retrained Space Lens Detector scored AUROC = 0.84, which shows the model has the minimum capacity to distinguish lenses from a different distribution. If we train the network again with 200 samples from the space-based dataset (1% of the space training set), the AUROC improves by 0.88. With 400 samples (2%), AUROC becomes 0.89, and with 600 samples (3%), AUROC improves to **0.902**. Each retraining was done independently of the others. The ROC curve of the Space Lens Detector trained with *r* band and was tested on the space-based dataset, and the improved ROC curves of the retrained model are shown in Fig. 3.13. The capacity of the model to identify the lenses improves if we train the model with very few samples from a different distribution, which indicates the adaptability of the encoder model in the presence of new data distributions. This feature also shows that the encoder models trained on simulated data can be optimised to detect strong lenses from real data using even a small sample of detected lenses.

Even though the encoder model performs better than the convolution models and the other models that participated in the challenge, the encoder models that have been trained here have a slight gap with a perfect classifier. We carefully examined the frequent false positives and the false negatives reported by various encoder models. Some of the images that have been identified as false positives and false negatives are given in Fig. 3.14 and

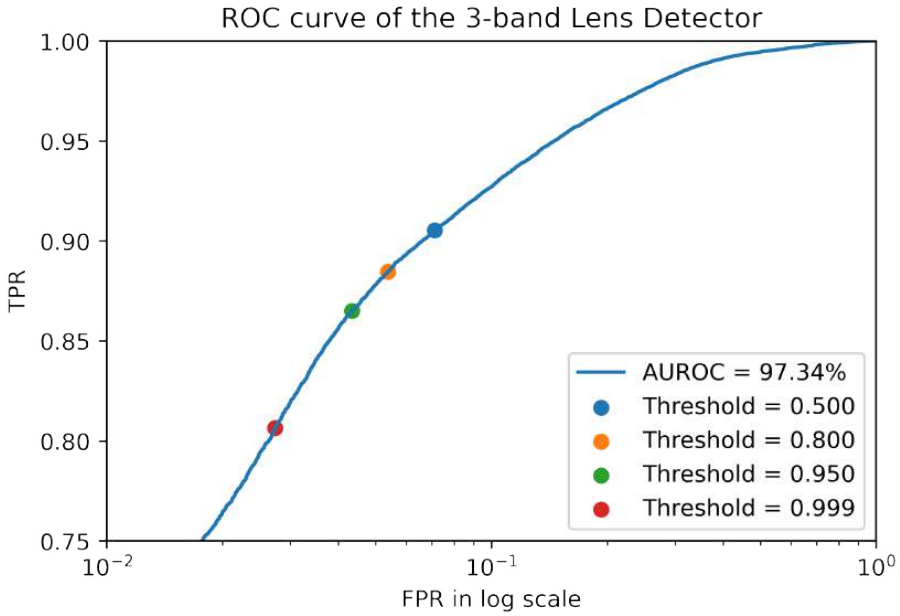


FIGURE 3.11: ROC curve of the 3-band Lens Detector.

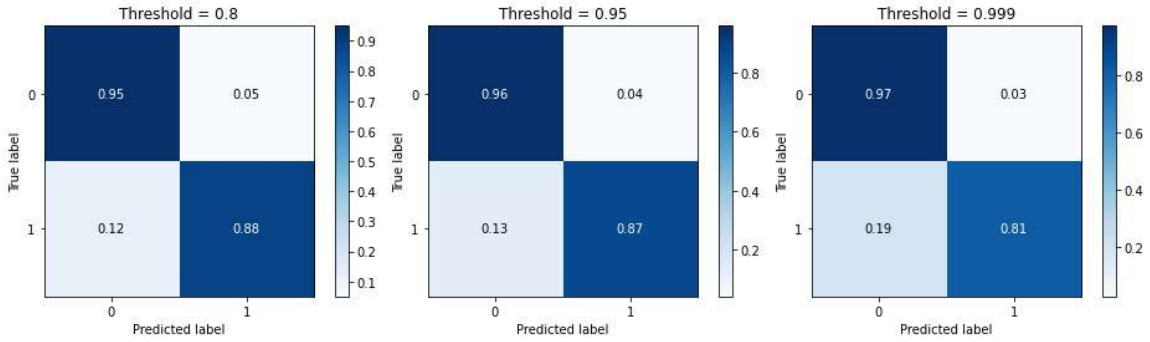


FIGURE 3.12: Confusion matrix of the 3-band Lens Detector on the challenge dataset for various thresholds. Class 0 represents the non-lensed images, and Class 1 represents the lensed images. The lower right represents TP in each confusion matrix, the lower left represents FN, the upper left represents TN, and the upper right represents FP.

Fig. 3.15. Looking at these false positives and false negatives, we can see that the encoder models are trying to find whether the input image has an arc-like structure or multiple distorted images. If we suppose the input image has any of these characteristics in at least one of the bands, the detector identifies the image as a strong lens. Similarly, if both these features are missing, then the detector classifies the image as a non-lens. In order to improve the performance of the models, we need the model to be trained on more realistic and complex data.

Since strong gravitational lensing is a rare phenomenon, we need to have a closer look at the cases of false negatives. In real surveys false positives can be filtered out from a candidate sample created by an automated classifier, whereas missing a true SL is not favoured. Looking at the relations between the parameters of the strong lens and the model's performance is a possible way to search for a space of parameters where it is more difficult to find lenses. We sorted the test dataset into ascending order in Einstein radii and divided it into three subclasses depending on the Einstein radius. The first quartile (25,000)

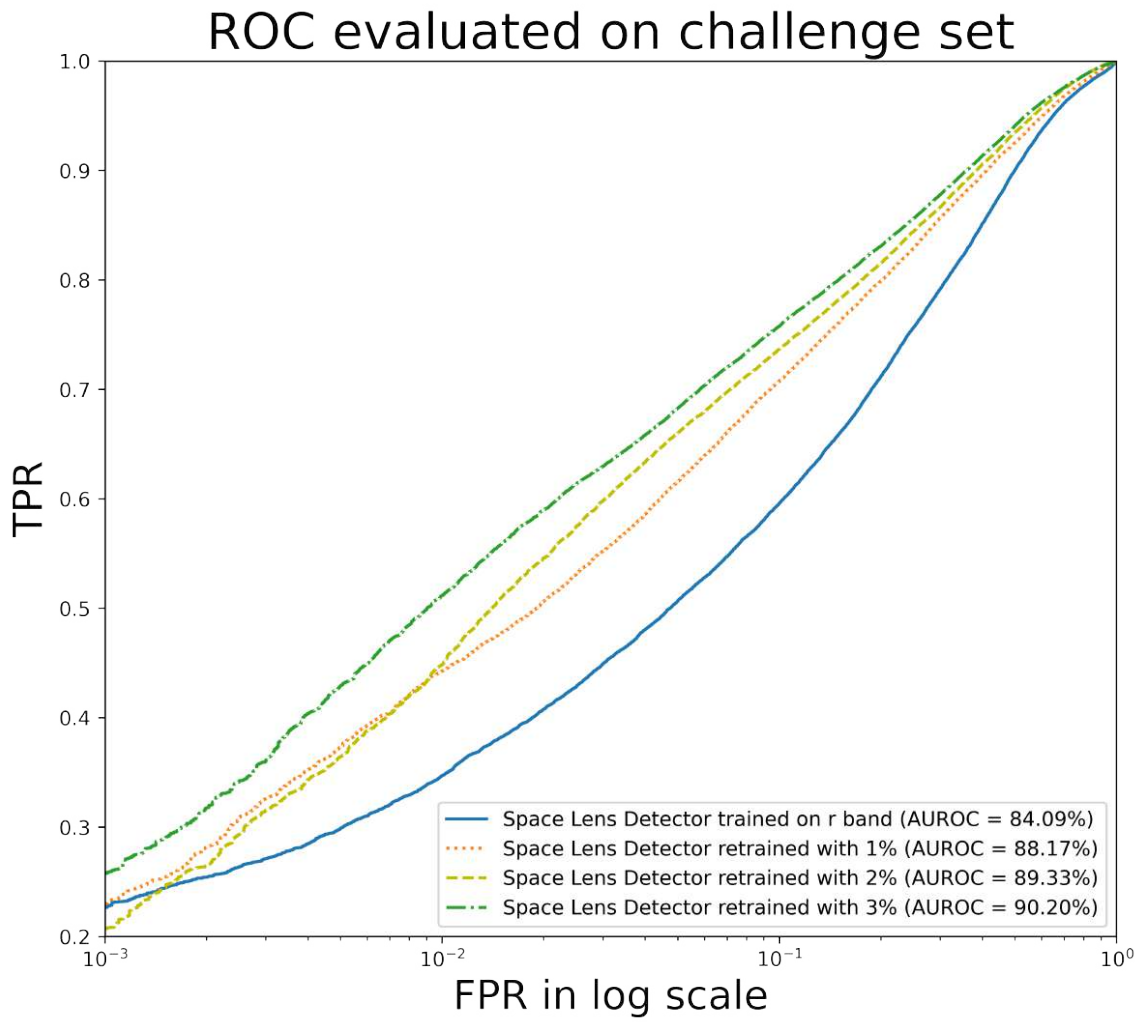


FIGURE 3.13: ROC curve of Space Lens Detector trained on *r* band and tested on the space-based dataset. Improved ROC curves of the model retrained using 200, 400, and 600 samples from the space dataset are also plotted.

and the third quartile (75,000) were used as the subclass boundaries. Looking at the confusion matrices (threshold = 0.8) from Fig. 3.16, we can see that the best performance of Lens Detector 15 is for the intermediate bin (0.873–3.547 arcsec). The performance of Lens Detector 15 is low for the first bin (0.3011–0.873 arcsec) and the third bin (3.547–10.08 arcsec). This means that Lens Detector 15 has difficulty finding SLs with small and large Einstein radii. A similar result for CNNs has also been reported by Li et al. (2020); Cañameras et al. (2020).

Analysing the parameters of the false negatives, we found that another important parameter that impacts the identification of strong lenses is the ratio of the flux in lensed pixels to the total flux. For a probability threshold of 0.8, we had 4981 false negatives in the Bologna Ground-Based Challenge. All of them had a very low ratio of the flux in lensed pixels to the total flux. Out of 4981 false negative cases, 4775 cases had a flux ratio of the source to the lens lower than 0.1. Similarly, 3667 out of 4981 cases in the sample of false negatives had a flux ratio of the source to the lens lower than 0.01. The Bologna Ground-Based Challenge dataset contained 10,818 true strong lenses with a flux ratio of the source to the lens lower than 0.01. Thus, Lens Detector 15 could identify 66% of the

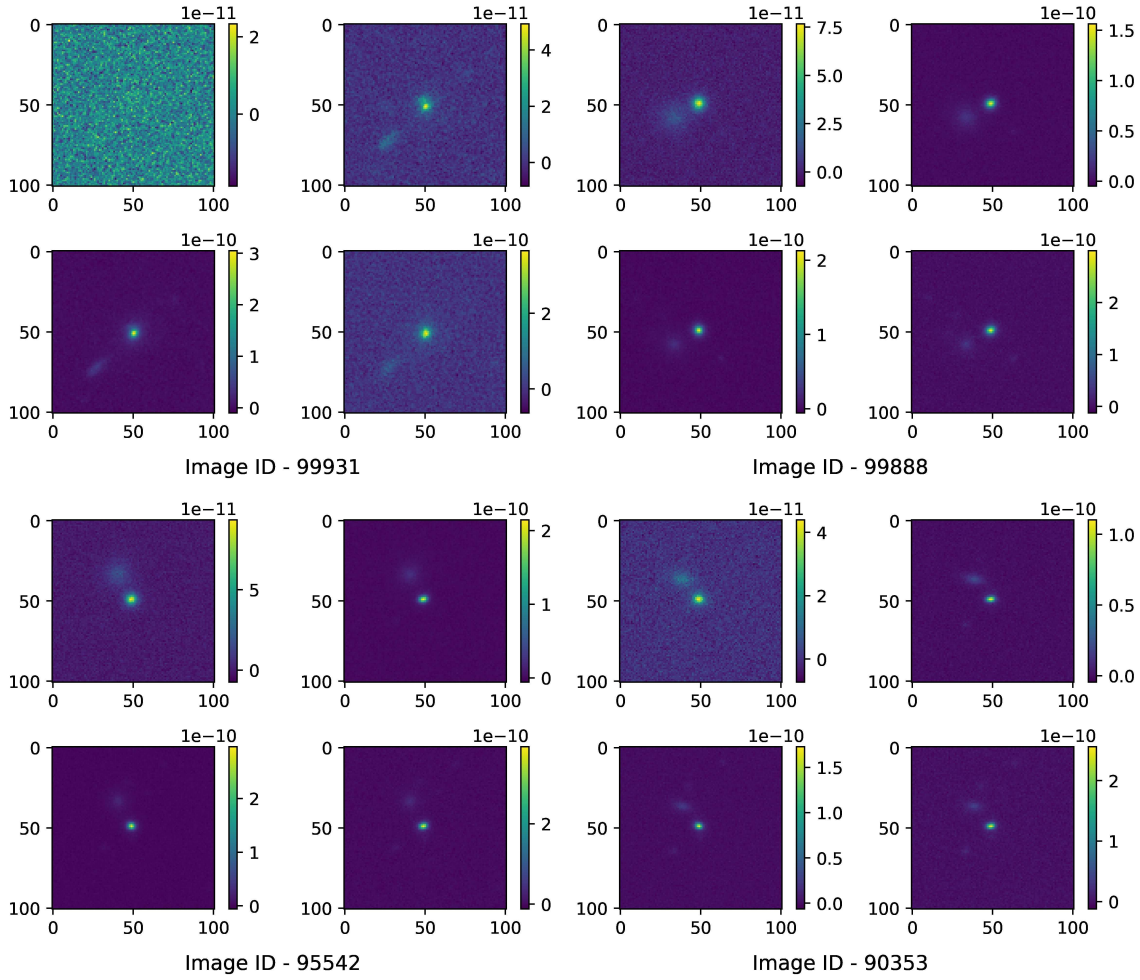


FIGURE 3.14: Four examples of false positives found by the encoder models. The channels shown are u (top left), g (top right), r (bottom left), and i (bottom right). Image ID from the test data is given below each set of images.

true strong lenses with a very low flux ratio (lower than 0.01). These results indicate that Lens Detector 15 will have trouble distinguishing strong lens candidates for a very low flux ratio between the source and the lens. The reason is that the distortions of the source galaxy due to lensing may not be bright enough to be detected by the models. This result is similar to the results reported by Li et al. (2020); Cañameras et al. (2020) where the CNN models have low performance on fainter SLs samples. The distribution of false negatives as a function of the ratio of flux in the lensed pixels to the total flux is plotted in Fig. 3.17.

We also would like to comment on another reported CNN model on the Bologna Lens Challenge, which is the LensCNN, achieving a total accuracy of 0.8749 (TP 0.8817 and TN 0.8682). It is the only CNN model where classification accuracy for the Bologna Lens challenge has been reported (Davies et al. 2019). Our results show that all of our encoder models have surpassed the LensCNN in total accuracy. Furthermore, the LensCNN model has also reported an AUROC of 0.96 on the challenge data, which is exceeded by most of our encoder models. In this context, it is worth mentioning that the LensCNN had approximately 10 $\times 10^6$ parameters. Lens Detectors outperformed the LensCNN with just 3 $\times 10^6$ of the parameters.

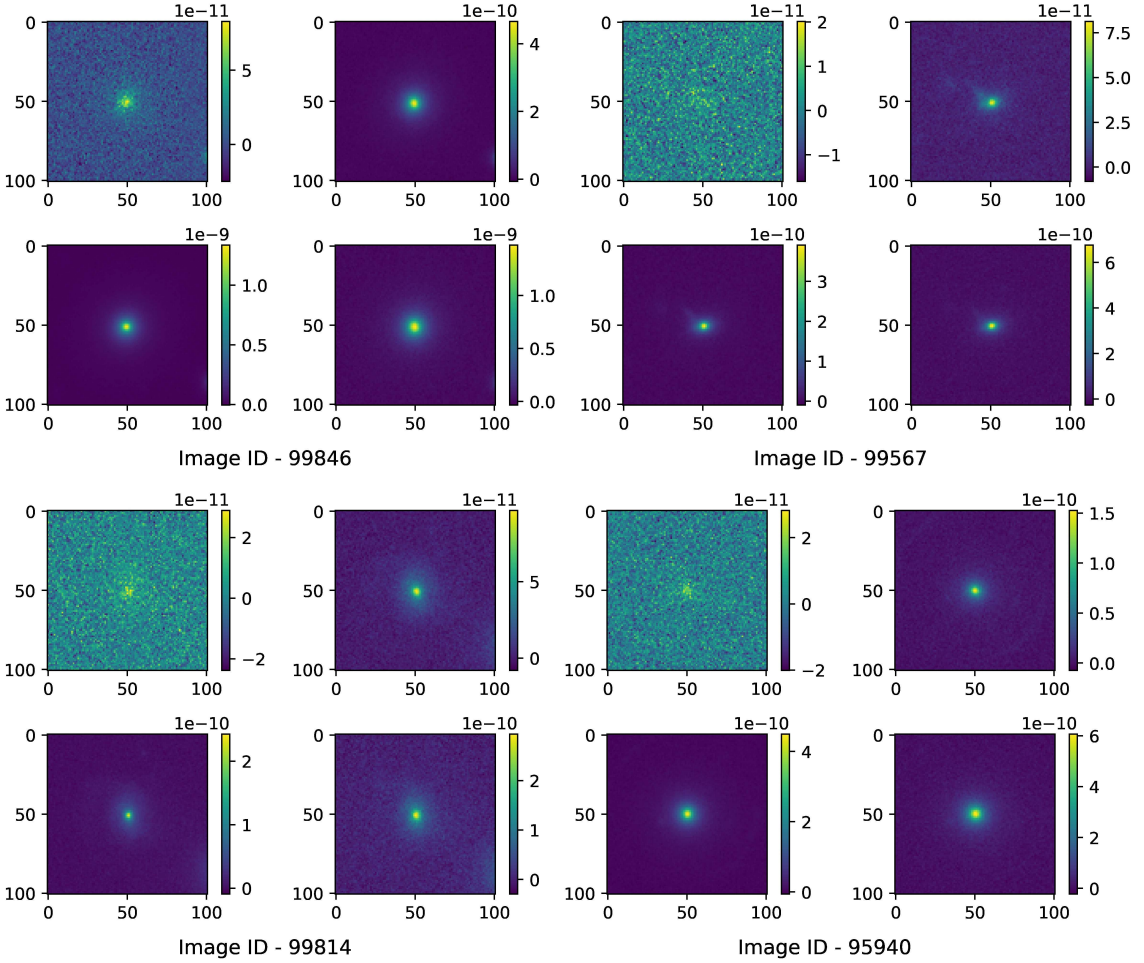


FIGURE 3.15: Four examples of false negatives found by the encoder models. The channels shown are u (top left), g (top right), r (bottom left), and i (bottom right). Image ID from the test data is given below each set of images.

3.5.4 Performance on real data

Since all our models have been trained on the simulated dataset provided by the Bologna Lens Challenge, it is critical to check if the trained model can identify strong lenses from real data. Ideally, we expect the encoder models to learn the general features of the strong lenses from the simulated data and to retrieve the potential lens candidates from the real data. Recently, Petrillo et al. (2019) have trained a three-band CNN (g, r, and i bands) and a single-band CNN (r band) on the data simulated using the real images from the KiDS survey and applied it to the KiDS DR4 data to identify potential strong lens candidates. To obtain a reasonable number of true positives and so as not to contaminate the sample with a large number of false positives, they chose 0.8 as the threshold for identifying a candidate as a strong lens for each CNN. Using these criteria, they shortlisted 3500 cases as potential strong lenses, and Petrillo et al. (2019) conducted a visual inspection to validate these candidates.

The potential candidates were classified into three classes, and each class was assigned a numerical score: Sure lens, 10 points; Maybe lens, 4 points; No lens, 0 points. As a result, the highest score that any candidate can obtain is 70, when all human classifiers think it is undoubtedly a lens. Using these criteria, Petrillo et al. (2019) shortlisted 1983 potential

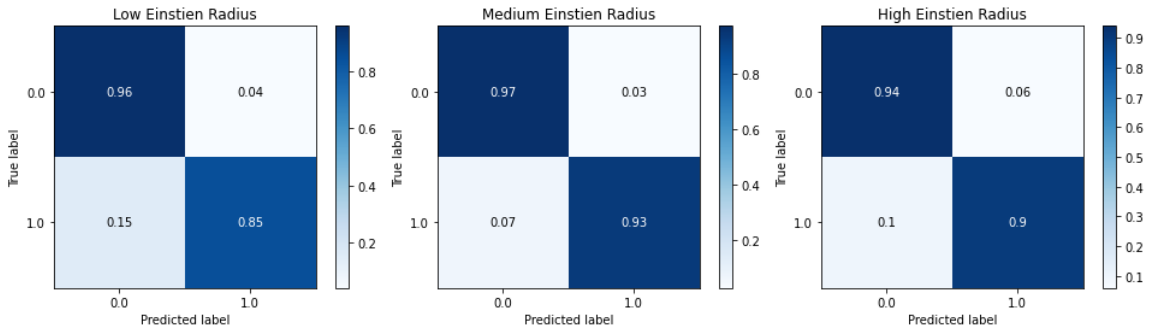


FIGURE 3.16: Confusion matrix of Lens Detector 15 with 0.8 as the threshold plotted for small Einstein radius (0.3011 – 0.873 arcsec), medium Einstein radius (0.873 – 3.547 arcsec) and high Einstein radius (3.547 – 10.08 arcsec). In the confusion matrix, the lower right represents TP, the lower left represents FN, the upper left represents TN, and the upper right represents FP.

strong lens candidates from the data selected by the two CNNs. The FITS files, probability scores of the CNNs reported in Petrillo et al. (2019), and numerical scores of visual inspection for each candidate are available publicly [3], and we chose this dataset to study the performance of the encoder model on real data.

Since the 1983 lens candidates were found together using a single-band CNN and a three-band CNN, some of the candidates found by the single-band CNN were not detected by the three-band CNN. Specifically, 946 candidates were missed by the three-band CNN (which means they were assigned a probability of less than 0.8) in the final sample of 1983 candidates. Similarly, 526 candidates identified by the three-band CNN were missed by the single-band CNN. To analyse the performance of the encoder model, we used the 3-band Lens Detector and tested it on the lens candidates found by Petrillo et al. (2019). After evaluating the model on real data, we created three classes using the visual inspection score as the reference. Class 0 sources have a low probability of being a lens (score less than 20 out of 70, and predictive value less than 0.8). Class 1 sources have an intermediate probability of being a lens (score between 20 and 50, and predictive value between 0.8 and 0.95). We are highly confident that Class 2 sources are strong lenses (score greater than 50, and a predictive value greater than 0.95). Using the probability scores predicted by the three-band CNN and using the visual scores, we plotted the confusion matrix for the three-band CNN along with the 3-band Lens Detector, which is given in Fig. 3.18.

Looking at the confusion matrix in Fig. 3.18, we can see that the encoder model can classify low (Class 0) and high (Class 2) probability cases similarly to a human expert. However, for class 1, which represents the cases with an intermediate probability of being a strong lens, the 3-band Lens Detector performs poorly. This is to be expected since the 3-band Lens Detector is trained as a binary classifier, and from Fig. 3.10 it is clear that the Lens Detector tends to assign very high or very low probability scores. Since there are three classes and the number of samples in each class is different, in order to compare the three-band CNN and the 3-band Lens Detector, we can calculate the weighted f_1 score by taking the mean of all per class f_1 scores while considering each class's support. Here, support refers to the number of actual occurrences of the class in the dataset. Using the visual scores as a reference, the weighted f_1 score of the three-band CNN is 0.601, and the weighted f_1 score of the 3-band Lens Detector is 0.822, which indicates that the 3-band Lens Detector is performing similarly to a human visual expert on the shortlisted SL candidates. Here we

[3] https://www.astro.rug.nl/lensesinkids/

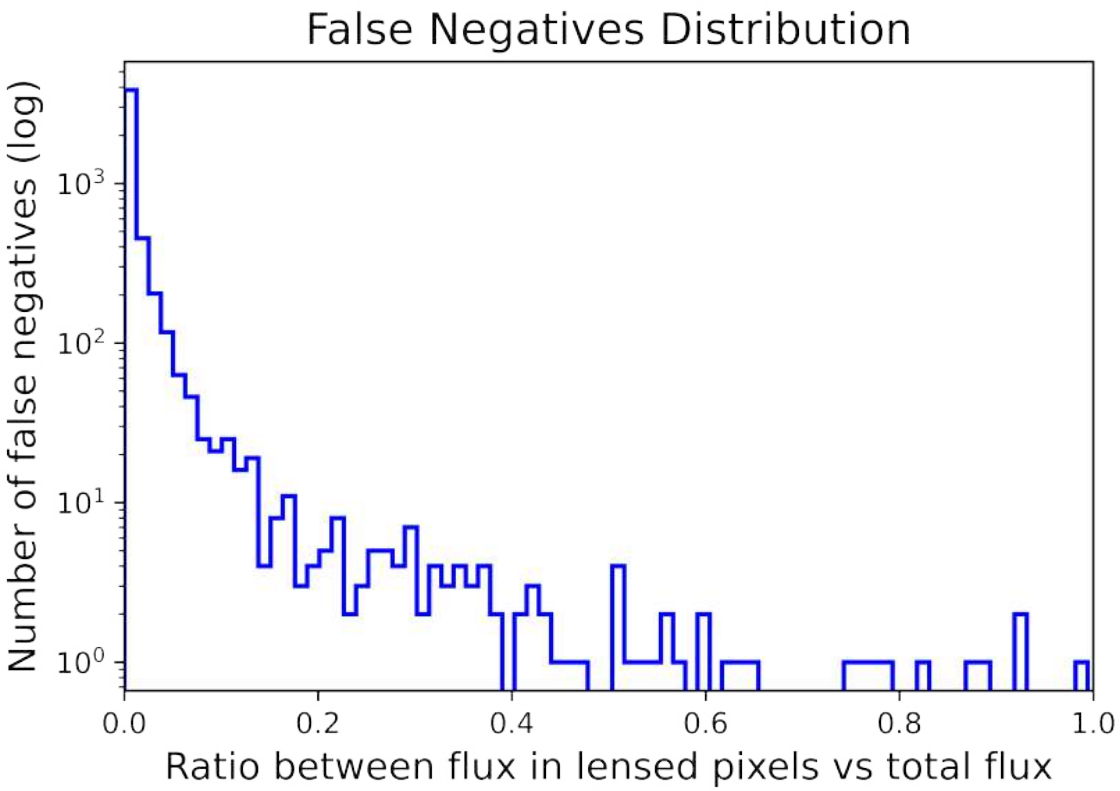


FIGURE 3.17: Distribution of false negatives as a function of the ratio of flux in the lensed pixels to the total flux.

have assumed that the output probabilities assigned by the three-band CNN and the team of visual experts are independent.

With these results, we cannot claim that the 3-band Lens Detector is better than the three-band CNN presented in Petrillo et al. (2019) or vice versa since we are testing the model on the already shortlisted candidates by the three-band CNN. However, we can claim that self-attention-based encoder models can detect strong lenses from real data in competition with CNNs. Another factor to be noted here is that the 3-band Lens Detector was trained on a complete data distribution compared to the training set of the three-band CNN, which was derived from the actual KiDS DR4 data. As mentioned earlier, the data in the Bologna Lens Challenge used KiDS as a reference, and they did not strictly resemble the data from KiDS. Thus, the data used in the Bologna Lens Challenge have a different data distribution compared to the KiDS DR4 data. Hence, if we were to directly apply our model to the data from KiDS DR4 without any fine-tuning, we are expected to find a lot of false positives. However, we show that the encoder models can adapt to different data distributions and improve their performance if we retrain, even with a small sample set from the new data distribution. Thus, the performance of the encoder model can be significantly improved if one retrains the 3-band Lens Detector with the data derived from the KiDS DR4 data. The fie-tuned models presented here in this work are used to find SGLs from KiDS DR4 by Grespan et al. (2024) as a followup of this analysis.

3.6 Conclusions

We have presented a novel machine learning approach known as the self-attention-based encoders to detect strong gravitational lenses. We have explored this new architecture's

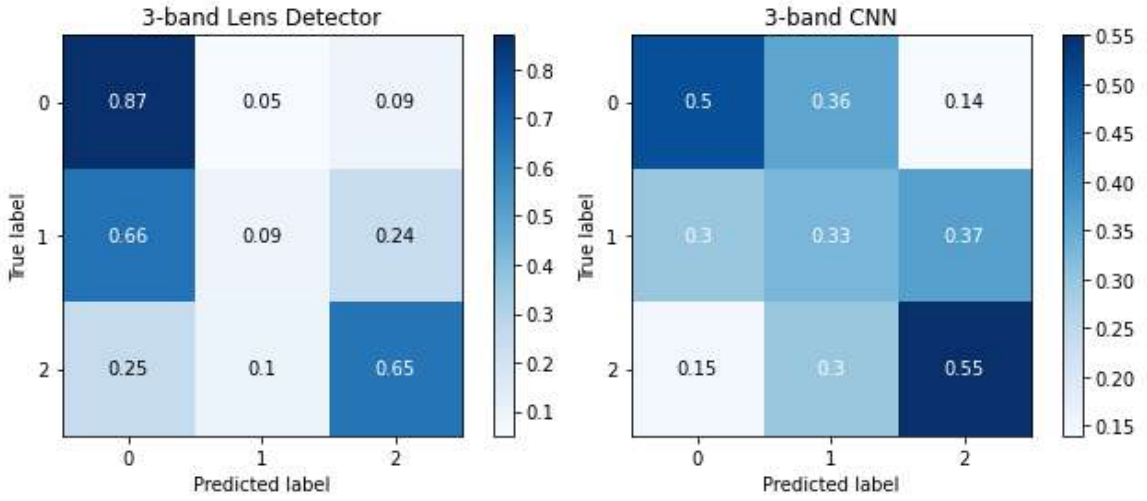


FIGURE 3.18: Confusion matrix of the 3-band Lens Detector and three-band CNN for the classification of the 1983 KiDS DR4 lens candidates (Petrillo et al. 2019). The candidates are classified into three categories based on the visual inspection scores by Petrillo et al. (2019), which is treated as the correct classification. The candidates are also classified into three categories based on the probability values generated by the 3-band Lens Detector introduced in this work and three-band CNN from Petrillo et al. (2019). The overlap between these classifications is shown in the confusion matrices.

possibilities to understand better how to apply the transformer models for image analysis using the data from the Bologna Lens challenge. Currently, most of the automated techniques employed to find strong lenses are based on CNNs. However, as noted by Metcalf et al. (2019), CNNs are prone to overfitting the training set. Here we showed that the self-attention-based architectures provide better stability and are less likely to overfit than CNNs. Another advantage of a self-attention-based encoder over a CNN is that it performs better with a fewer number of trainable parameters. Hence, self-attention-based encoder models can be considered a better alternative to CNNs and other automated methods.

Here we have described the 21 encoder models we created to study the application of self-attention-based models for SL detection using the data from the Bologna Lens Challenge. We have presented the three best encoder models, which provide more reliable performance than those participating in the Bologna Lens Challenge. Lens Detector 21 scored an AUROC of **0.9809**, which is equivalent to the top AUROC achieved in the challenge. Similarly, Lens Detector 16 scored a TPR_0 **0.225** higher than any model that participated in the challenge, and surpassed the top TPR_0 (0.09) achieved by the CNNs by a high margin. We consider Lens Detector 15 to be the best encoder model as it scored 0.14 and 0.48 respectively for TPR_0 and TPR_{10}, outperforming the CNN models to a greater extent and also scoring an AUROC of 0.9783, which is very close to the top AUROC.

From our analysis, we were able to point out that the encoder models have more stability than CNNs, which minimises the need for human interaction or monitoring. Similarly, the encoder models were better than the CNN models in classifying lenses and non-lenses by assigning high probability scores for the lens ($p \approx 1$) and non-lens ($p \approx 0$) systems. In addition, the architecture we proposed here is very simple and robust and has a high resistance to overfitting. We could train models as deep as 25 layers and for 2000 epochs without any sign of overfitting. With a simple 8-layer deep CNN, we were able to surpass the performance of a 46-layer deep RNN and surpass all the other models to a great extent.

We tested our model on the 1983 potential strong lens candidates from the KiDS DR4

data found in Petrillo et al. (2019). We were able to closely mimic a human visual expert in identifying the strong lenses. Even though we cannot claim to outperform the CNN model presented in Petrillo et al. (2019), we confirm that the encoder models can perform well on the real data. Since we have tested the network on a different data distribution than it was trained on, we expect to improve the performance of the encoder model if the training and testing data distribution are similar. In the future, we are planning to train the encoder models on more complex data derived from real data and test them on real data to find more potential lens candidates. Even though we have glimpsed at the adaptability of the encoder models to different data distributions, further studies are needed to establish the full scope of this architecture.

In the upcoming era of big data in astronomy, automated methods are expected to play a crucial role. Better and alternative automated methods have to be consistently investigated to advance the scientific study in this scenario. From our study it is clear that the search for strong lenses in the current and upcoming wide-field surveys such as KiDS (Kuijken et al. 2019), HSC (Aihara et al. 2019), LSST (Ivezić et al. 2019; Verma et al. 2019), Euclid (Euclid Collaboration et al. 2022b), and WFIRST (Koekemoer 2019) can be achieved using self-attention-based encoder models with better performance compared to CNNs.

4

Shedding light on low-surface-brightness galaxies in dark energy surveys with transformer models

This chapter originally appeared as 'Shedding light on low-surface-brightness galaxies in dark energy surveys with transformer models' by Thuruthipilly, H. et al. 2024, Astronomy & Astrophysics, (Thuruthipilly et al. 2024b).

4.1 Introduction

Low-surface-brightness galaxies (LSBGs) are most often defined as galaxies with a central surface brightness fainter than the night sky or galaxies with B-band central surface brightness $\mu_0(B)$ below a certain threshold value. In literature, the threshold values of $\mu_0(B)$ for classifying a galaxy as LSBG vary among different works, ranging from $\mu_0(B) \geq 23.0$ mag arcsec^{-2} (Bothun et al. 1997) to $\mu_0(B) \geq 22.0$ mag arcsec^{-2} (Burkholder et al. 2001).

It is estimated that the LSBGs only contribute a few percentages ($< 10\%$) to the local luminosity and stellar mass density of the observable universe (Bernstein et al. 1995; Driver 1999; Hayward et al. 2005; Martin et al. 2019). However, LSBGs are considered to account for a significant fraction ($30\% \sim 60\%$) of the total number density of galaxies (McGaugh 1996; Bothun et al. 1997; O'Neil & Bothun 2000; Haberzettl et al. 2007; Martin et al. 2019), and as much as 15% of the dynamical mass content of the universe (Driver 1999; Minchin et al. 2004). These numbers imply that LSBGs can contribute significantly to our understanding of the physics of galaxy evolution and cosmological models. However, as their name indicates, LSBGs are very faint systems, and due to the observational challenges in detecting them, LSBGs remain mostly an unexplored realm.

In recent years, despite the observational challenges, advances in digital imaging have improved our ability to detect LSBGs. The first known and the largest LSBG to be identified and verified is Malin 1, serendipitously discovered by Bothun et al. (1987) during a survey of galaxies of low surface brightness in the Virgo cluster. Notably, Malin 1 is the largest spiral galaxy known until today (e.g., Impey et al. 1988; Junais et al. 2020; Galaz et al. 2022). Current searches for LSBGs have shown that they exhibit a wide range of physical sizes (E Greene et al. 2022) and can be found in various types of environments, ranging from satellites of local nearby galaxies (Danieli et al. 2017; Cohen et al. 2018), ultra-faint

satellites of the Milky Way (McConnachie 2012; Simon 2019), galaxies found in the field (Leisman et al. 2017; Prole et al. 2021), to members of massive galaxy clusters like Virgo (Mihos et al. 2015, 2017; Junais et al. 2022) and Coma (van Dokkum et al. 2015a; Koda et al. 2015).

LSBGs also consist of several sub-classes based on their physical size, surface brightness and gas content. Ultra-diffuse galaxies (UDGs) represent a subclass of LSBGs characterized by their considerable size, comparable to that of Milky Way-like galaxies, yet exhibiting very faint luminosities akin to dwarf galaxies. Although the term 'UDG' was coined by van Dokkum et al. (2015a), such galaxies were identified in several earlier studies in the literature (Sandage & Binggeli 1984; McGaugh & Bothun 1994; Dalcanton et al. 1997; Conselice et al. 2003a). Similarly, giant LSBGs (GLSBGs) form another sub-class of LSBGs that are extremely gas-rich ($M_{HI} > 10^{10}\ M_{\odot}$), faint and extended (Sprayberry et al. 1995; Saburova et al. 2023). The formation and evolution of extreme classes like UDGs and GLSBGs are still debated (Amorisco & Loeb 2016a; Di Cintio et al. 2017; Saburova et al. 2021; Benavides et al. 2023; Laudato & Salzano 2023).

To comprehend the formation of various types LSBGs in different environments,studying them extensively across different environments (galaxy clusters vs field) over a large area of the sky is crucial. Recently, Greco et al. (2018) detected 781 LSBGs in the Hyper Suprime-Cam Subaru Strategic Program (HSC SSP) in a blind search covering around 200 deg^2 of the sky from the Wide layer of the HSC SSP. Similarly, in a recent study, Tanoglidis et al. (2021b) utilised a support vector machine (SVM) and visual inspection to analyse the first three years of data from the Dark Energy Survey (DES). They identified more than 20 000 LSBGs, which is currently the largest LSBG catalogue available.

A common feature observed in both of these untargeted searches for LSBGs was the significant presence of low-surface brightness artefacts. As pointed out in Tanoglidis et al. (2021b), these artefacts predominantly consist of diffuse light from nearby bright objects, galactic cirrus, star-forming tails of spiral arms and tidal streams. These artefacts typically pass the simple selection cuts based on photometric measurements and often make up the majority of the LSBG candidate sample. These contaminants need to be removed, which is often accomplished using semi-automated methods with a low success rate and visual inspection, which is more precise but time-consuming.

For example, in HSC SSP, Greco et al. (2018) applied selection cuts on the photometric measurements from `SourceExtractor` (Bertin & Arnouts 1996). This led to the selection of 20 838 LSBG candidates. Using a galaxy modelling pipeline based on `imfit` (Erwin 2015), the sample size was subsequently reduced to 1 521. However, after the visual inspection, only 781 candidates were considered confident LSBGs, which is around 4% of the preliminary candidate sample and 50% of the sample selected by the pipeline. Similarly, in DES, Tanoglidis et al. (2021b) shortlisted 419 895 LSBG candidates using the selection cuts on `SourceExtractor` photometric measurements. After applying a feature-based machine learning (ML) classification (SVM) on the photometric measurements, the candidate sample was further reduced to 44 979 objects. However, a significant number of false positives still remained, and only 23 790 were later classified as confident LSBGs. Therefore, these numbers indicate that the occurrence of LSBGs in these methods is roughly 5% for the initial selection and 50% for the subsequent selection.

The upcoming large-scale surveys, such as Legacy Survey of Space and Time (LSST; Ivezić et al. 2019) and Euclid (Euclid Collaboration et al. 2022b), are expected to observe billions of astronomical objects. In this scenario, relying solely on photometric selection cuts or semi-automated methods such as galaxy model fitting would not be practical to identify LSBGs confidently. Furthermore, the accuracy of the classification methodology between LSBGs and artefacts must be improved to achieve meaningful results. Hence, this

situation demands more effective and efficient automation methodologies for the searches of LSBGs.

Recently, the advancements in deep learning have opened up a plethora of opportunities and have been widely applied in astronomy. Particularly for analysing astronomical images, convolutional neural networks (CNNs) have emerged as a state-of-the-art technique. For example, the CNNs have been used for galaxy classification (Pérez-Carrasco et al. 2019), galaxy merger identification (Pearson et al. 2022), supernova classification (Cabrera-Vives et al. 2017) and finding strong gravitational lenses (Schaefer et al. 2018; Davies et al. 2019; Rojas et al. 2022). One of the fascinating features of CNNs is their ability to directly process the image as input and learn the image features, making them one of the most popular and robust architectures in use today. Generally, the learning capacity of a neural network increases with the number of layers in the network. The first layers of the network learn the low-level features, and the last layers learn more complex features (Russakovsky et al. 2015; Simonyan & Zisserman 2015).

One of the main requirements for creating a trained CNN is a sufficiently large training dataset that can generalise the features of the data we are trying to analyse. Recently, Tanoglidis et al. (2021a) utilised a catalogue of over 20 000 LSBGs from DES to classify LSBGs from artefacts using a CNN for the first time and achieved an accuracy of 92% and a true positive rate of 94%.

While CNNs have been the dominant choice for analysing image data in astronomy, the current state-of-the-art models for computer vision are transformers. Transformers were initially introduced in natural language processing (NLP) as an attention-based model (Vaswani et al. 2017). The fundamental concept behind the transformer architecture is the attention mechanism, which has also found a broad range of applications in machine learning (Zhang et al. 2019; Fu et al. 2019; Parmar et al. 2019; Zhao et al. 2020; Tan et al. 2021). In the case of NLP, attention calculates the correlation of different positions of a single sequence to calculate a representation of the sequence. Later the idea was adapted to computer vision and has been used to produce state-of-the-art models for various image processing tasks like image classification (Wortsman et al. 2022), and image segmentation (Chen et al. 2023).

Generally, two categories of transformers are present in the literature. The first type integrates both CNNs and attention to perform the analysis. An example of this type is the Detection Transformer (DETR) proposed for end-to-end object detection by Carion et al. (2020). The key idea behind using CNNs and Transformers together is to leverage the strengths of both architectures. CNNs excel at local feature extraction, capturing low-level details, and spatial hierarchies, while attention layers excel at modelling global context and long-range dependencies. The second class of transformers is the models that do not use a CNN and operate entirely based on self-attention mechanisms. An example of this type is the Vision transformer (ViT) proposed for object classification by Dosovitskiy et al. (2021). ViTs have demonstrated remarkable performance in image classification tasks and have surpassed the accuracy of CNN-based models on various benchmark datasets (Dosovitskiy et al. 2021; Yu et al. 2022; Wortsman et al. 2022).

Even though transformers have been introduced very recently in astronomy, they have already found a wide variety of applications. For example, the transformer models have been used to detect and analyse strong gravitational lensing systems (Thuruthipilly et al. 2022, 2024a; Huang et al. 2023; Jia et al. 2023), representing light curves which can be used further for classification or regression (Allam & McEwen 2021), and classifying multi-band light curves of different supernovae (SN) types (Pimentel et al. 2023).

In this paper, we explore the possibilities of transformers in classifying LSBGs from artefacts in DES and compare the performance of transformers with the CNNs presented in Tanoglidis et al. (2021a). We also use the transformer models to look for new LSBGs that

the previous searches may have missed. For comparison purposes, throughout this work, we follow the LSBG definition from Tanoglidis et al. (2021b), based on the g-band mean surface brightness ($\bar{\mu}_{eff}$) and the half-light radii ($r_{1/2}$). We consider LSBGs as galaxies with $\bar{\mu}_{eff} > 24.2$mag arcsec^{-2} and $r_{1/2} > 2.5''$.

The paper is organised as follows. Sect. 4.2 discusses the data we used to train our models and look for new LSBGs. Sect. 4.3 provides a brief overview of the methodology used in our study, including the models' architecture, information on how the models were trained, and the details about the visual inspection. The results of our analysis are presented in Sect. 4.4. A detailed discussion of our results and the properties of the newly identified LSBGs are analysed in Sect. 4.5 and Sect. 4.6 respectively. Further analysis of the clustering of LSBGs is presented in Sect. 4.7 and a detailed discussion on the UDGs, which are identified as a subsample of LSBGs, is presented in Sect. 4.8. Sect. 4.9 concludes our analysis by highlighting the significance of LSBGs, the impact of methodology in finding LSBGs, and the future prospect with LSST.

4.2 Data

4.2.1 Dark Energy Survey

The Dark Energy Survey (DES; Abbott et al. 2018a, 2021) is a six-year observing program (2013-2019) covering $\sim$ 5000 deg^2 of the southern Galactic cap in the optical and near-infrared regime using the Dark Energy Camera (DECam) on the 4-m Blanco Telescope at the Cerro Tololo Inter-American Observatory (CTIO). The DECam focal plane comprises 62 2k $\times$ 4k charge-coupled devices (CCDs) dedicated to science imaging and 12 2k $\times$ 2k CCDs for guiding, focus, and alignment. The DECam field-of-view covers 3 deg^2 with a central pixel scale of 0.263 arcsec pixel^{-1} (Flaugher et al. 2015). To address the gaps between CCDs, DES utilises a dithered exposure pattern (Neilsen et al. 2019) and combines the resulting individual exposures to form coadded images, which have dimensions of 0.73 $\times$ 0.73 degrees (Morganson et al. 2018). The DES has observed the sky in $grizY$ photometric bands with approximately 10 overlapping dithered exposures in each filter (90 sec in $griz$-bands and 45 sec in Y-band).

4.2.2 DES DR1 and the gold catalogue

In this work, we use the image data from the dark energy survey data release 1 (DES DR1; Abbott et al. 2018a) and the DES Y3 gold coadd object catalogue (DES Y3_gold_2_2.1) obtained from the first three years of the DES observations (Sevilla-Noarbe et al. 2021). The DES DR1 comprises optical and near-infrared imaging captured over 345 different nights between August 2013 and February 2016. The median 3σ surface brightness limits of the g, r, and i-bands of DES DR1 are 28.26, 27.86, and 27.37 mag arcsec^{-2}, respectively (Tanoglidis et al. 2021b). It is worth mentioning that the DES source detection pipeline has not been optimised for detecting large, low surface-brightness objects (Morganson et al. 2018). Thus, the above-mentioned surface brightness values can be considered as the limits for detecting faint objects in each band. The gold catalogue shares the same single image processing, image coaddition, and object detection as the DES DR1. The objects in the gold catalogue were detected using `SourceExtractor` (Bertin & Arnouts 1996) and have undergone selection cuts on minimal image depth and quality, additional calibration, and deblending. The median coadd magnitude limit of the DES Y3 gold object catalogue at a signal-to-noise ratio (S/N) = 10 is $g = 24.3$ mag, $r = 24.0$ mag, and $i = 23.3$ mag (Sevilla-Noarbe et al. 2021). The gold catalogue contains around 319 million astronomical objects, which we used for searching LSBGs in DES. For a detailed review and discussion

on the data from the DES, please refer to Abbott et al. (2018a) and Sevilla-Noarbe et al. (2021).

We reduced the number of objects processed in our study using preselections similar to Greco et al. (2018) and Tanoglidis et al. (2021b). We first removed objects classified as point-like objects in the gold catalogue, based on the i-band `SourceExtractor SPREAD_MODEL` parameter and `EXTENDED_CLASS_COADD` as described in Tanoglidis et al. (2021b). In addition, we constrained the g-band half-light radius (`FLUX_RADIUS_G`) and surface brightness (`MUE_MEAN_MODEL_G`) within the range of $2.5'' < r_{1/2} < 20''$ and $24.2 < \bar{\mu}_{eff} < 28.8$ mag arcsec^{-2}, respectively. Furthermore, we also limited our sample to objects with colors (using the `MAG_AUTO` magnitudes) in the range:

$$-0.1 < g - i < 1.4, \tag{4.1}$$

$$(g - r) > 0.7 \times (g - i) - 0.4, \tag{4.2}$$

$$(g - r) < 0.7 \times (g - i) + 0.4. \tag{4.3}$$

These color cuts are based on Greco et al. (2018) and Tanoglidis et al. (2021b). As mentioned by Greco et al. (2018), these color requirements will remove the spurious detections due to optical artefacts detected in all bands and blends of high-redshift galaxies. Finally, we also restricted the axis ratio (`B_IMAGE/A_IMAGE`) of each object to be greater than 0.3 to remove the artefacts like the highly elliptical diffraction spikes. Our complete selection criteria were based on the selection criteria presented in Appendix B of Tanoglidis et al. (2021b). After the preliminary selections using the `SourceExtractor` parameters from the DES Y3 gold catalogue, our sample contains 419 784 objects.

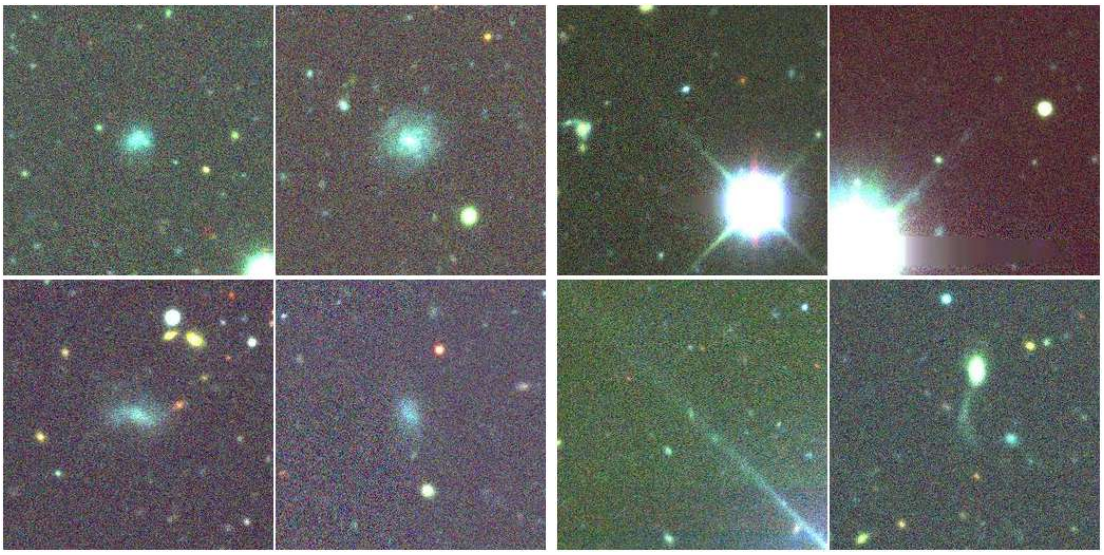
4.2.3 Training data

All of the trained, validated, and tested models in this study used the labelled dataset of LSBGs and artefacts identified from DES by Tanoglidis et al. (2021b). Below, we briefly summarise the primary steps taken by Tanoglidis et al. (2021b) in constructing the LSBG catalogue.

- The `SourceExtractor` parameters from the DES Y3 gold coadd object catalogue presented by Sevilla-Noarbe et al. (2021) were used to create the initial selection cuts, as discussed in Sect. 4.2.2.
- The candidate sample was further reduced using an SVM to classify artefacts and LSBGs. The SVM was trained with a manually labelled set of $\sim$ 8 000 objects (640 LSBGs) and using the `SourceExtractor` parameters as features for learning.
- From the candidate sample generated through SVM, over 20 000 artefacts were excluded upon visual inspection. Most of the rejected objects that had passed SVM's feature-based selection were found to be astronomical artefacts (such as galactic cirrus, star-forming extensions of spiral arms, and tidal streams) rather than instrumental artefacts (such as scattered light emitted by nearby bright objects) during visual inspection.
- Objects that passed the visual inspection were subjected to Sérsic model fitting and Galactic extinction correction. Following this, new selection cuts were applied to the updated parameters, and the final LSBG catalogue containing 23 790 LSBGs was created.

For training our classification models, we selected LSBGs from the LSBG catalogue as the positive class (label - 1) and the objects rejected in the third step (visual inspection) by

Tanoglidis et al. (2021b) as the negative class (label - 0). The catalogues for the positive and negative classes are publicly available, and we used these catalogues to create our training dataset[1]. The selection of the artefacts and LSBGs for training was random, and after selection, we had 18 474 artefacts and 23 103 LSBGs. However, when we further inspected these LSBGs and artefacts, we found that there were 797 objects belonging to both classes. After conducting a thorough visual examination, we identified that these are, in fact, LSBGs that had been mistakenly categorized as artefacts in the publicly accessible artefact catalogue. However, we avoided these 797 objects from our training set to avoid contamination and ambiguity among classes during training. We generated multi-band cutouts for each object in the flexible image transport system (FITS) format using the cutout service provided in the DES public data archive. Each cutout corresponds to a 67.32″ × 67.32″ (256 × 256 pixels) region of the sky and is centred at the coordinates of the object (LSBG or artefact). We resized the cutouts from their initial size to 64 × 64 pixels to reduce the computational cost. The cutouts of g, r and z-bands were stacked together to create the dataset for training the models. Examples of LSBGs and artefacts used for training the model are shown in Fig. 4.1. Our training catalogue contains 39 983 objects, out of which 22 306 are LSBGs and 17 677 are artefacts. Before training, we randomly split the full sample into a training set, a validation set and a test set, each consisting of 35 000, 2 500, and 2 483 objects, respectively.



(A) (B)

FIGURE 4.1: Four examples of LSBGs (4.1a) and artefacts (4.1b) used in the training data. Each image of the LSBG and artefact corresponds to a 67.32″ × 67.32″ region of the sky. Images were generated by combining the g, r and z bands using `APLpy` package (Robitaille & Bressert 2012).

4.3 Methodology

4.3.1 LSBG Detection Transformer (LSBG DETR)

We implemented four transformer models that use a CNN backbone and self-attention layers to classify the labels, which we call LSBG DETR (LSBG detection transformers) models in general. Each individual LSBG DETR model is followed by a number indicating their

[1] https://github.com/dtanoglidis/DeepShadows/blob/main/Datasets

chronological order of creation. The LSBG DETR models have an 8-layer CNN backbone to extract feature maps from the input image. The general architecture of the LSBG DETR models used here are presented in Sec. 2.5.2.

4.3.2 LSBG Vision

We have created four transformer models similar to the Vision Transformer (ViT) introduced by Google Brain (Dosovitskiy et al. 2021), which we call LSBG vision transformers (LSBG ViT) in general. Similar to LSBG DETR models, each individual model is followed by a number indicating their chronological order of creation. For a detailed discussion on ViT models, please refer to Dosovitskiy et al. (2021). The general architecture of the LSBG ViT models used here are presented in Sec. 2.5.3.

4.3.3 Training

All of the LSBG DETR and LSBG ViT models were trained with an initial learning rate of $\alpha = 10^{-4}$. We used the exponential linear unit (ELU) function as the activation function for all the layers in these models (Clevert et al. 2016). We initialise the weights of our model with the Xavier uniform initialiser (Glorot & Bengio 2010b), and all layers are trained from scratch by the ADAM optimiser with the default exponential decay rates (Kingma & Ba 2015). We have used the early stopping callback from `Keras` [2] to monitor the validation loss of the model and stop training once the loss was converged. The models LSBG DETR 1 and 4 had 8 heads and were trained for 150 and 93 epochs, respectively. Similarly, the LSBG DETR 2 and 3 had 12 heads and were trained for 134 and 105 epochs, respectively. Coming to the LSBGS ViT models, the hyperparameters we varied were the size of the image patches, the number of heads and the number of transformer encoder layers. The hyperparameters for the all the LSBG DETR models were customized based on the results from Thuruthipilly et al. (2022), which extensively investigated the hyperparameter configurations of DETR models. When it comes to the LSBG vision transformer models, we maintained the hyperparameters from the LSBG DETR models such as learning rate, and batch size, except for adjustments in image patch size, the count of attention heads, and the number of transformer encoder layers. We varied these parameters and the four best models are presented in Table 4.1. In the spirit of reproducible research, our code for LSBG DETR and LSBG ViT is publicly available [3].

TABLE 4.1: Table showing the name of the model, size of the image patches (s), number of heads (h), number of transformer encoder layers (T) and the number of epochs taken to train (e) the four vision models in chronological order of creation.

Model Name	s	h	T	e
LSBG VISION 1	4	12	4	55
LSBG VISION 2	4	12	8	55
LSBG VISION 3	6	12	4	67
LSBG VISION 4	6	16	8	67

[2] https://keras.io/api/callbacks
[3] https://github.com/hareesht23/

4.3.4 Ensemble Models

We had two classes of transformers (LSBG DETR and LSBG ViT) with four models in each class, and we used an ensemble model of these four models for each class to look for new LSBGs from DES DR1. Ensemble models in deep learning refer to combining multiple models to create a single model that performs better than the individual models. The idea behind ensemble models is to reduce the generalisation error and increase the stability of the system by taking into account multiple sources of information. Various kinds of ensemble learning exist in the literature, and they have been found helpful in a broad range of machine learning problems (Wang et al. 2022). For a detailed review of ensemble methods, please refer to Domingos & Hulten (1999) and Dietterich (2000). One of the easiest and most common ensemble methods is model averaging. In model averaging, multiple models are trained independently on the same training data, and the outputs of the models are averaged to make the final prediction. One of the main advantages of model averaging is that it is computationally efficient and does not require any additional training time. It also allows the use of different types of model architectures and can take advantage of their strengths and weaknesses and improve overall performance. Here we use averaging to create the ensemble models for LSBG DETR and LSBG ViT.

4.3.5 Sérsic fitting

The candidates identified independently by both LSBG DETR and LSBT ViT ensemble models were subjected to a single component Sérsic fitting using `Galfit` (Peng et al. 2002). This was done to re-estimate the $\bar{\mu}_{eff}$ and $r_{1/2}$ values of our LSBG candidates that we initially used for our sample selection. We employed a single-component Sérsic fitting method to align with the LSBG search methodology of Tanoglidis et al. (2021b), who also utilized a similar approach. However, we also note that Sérsic fitting does not always capture the full light from a galaxy.

We used the magnitude (`MAG_AUTO`) and radius (`FLUX_RADIUS`) values from the gold catalogue as an initial guess for the `Galfit` procedure. Moreover, the Sérsic index (n) and axis ratio (q) were initialised to be at a fixed value of 1 and were allowed to vary only within the range of $0.2 < n < 4.0$ and $0.3 < q \leq 1.0$, respectively. A similar fitting procedure was done for both the g-band and i-band images of our sample. After the fitting, we excluded all the sources with poor/failed fits with either a reduced $\chi^2 > 3$ or if their `Galfit` magnitude estimates diverge from their initial `MAG_AUTO` values by more than one mag. We have also excluded the cases where the estimated n and q values do not converge and are on the edge of the range specified above. For the remaining galaxies, we re-applied our g-band sample selection criteria of $\bar{\mu}_{eff} > 24.2$ mag arcsec^{-2} and $r_{1/2} > 2.5''$, following Tanoglidis et al. (2021b). The $\bar{\mu}_{eff}$ values were calculated using the relation given by Eq. 4.4:

$$\bar{\mu}_{eff} = m + 2.5 \times log_{10}(2\pi r_{1/2}^2), \tag{4.4}$$

where $\bar{\mu}_{eff}$ is the mean surface brightness within the effective radius, m is the total magnitude and $r_{1/2}$ is the half-light radius in a specific band estimated from `Galfit`. For all our measurements, we also applied a foreground Galactic extinction correction using the Schlegel et al. (1998) maps normalised by Schlafly & Finkbeiner (2011) and a Fitzpatrick (1999) dust extinction law.

4.3.6 Visual Inspection

Only the candidates identified independently by LSBG DETR and LSBT ViT ensemble models and passed the selection criteria for being an LSBG with the updated parameters from the `Galfit` were considered for visual inspection. This refined sample was subjected to visual inspection by two authors independently. Candidates identified as LSBG by both authors were treated as confident LSBGs, and candidates identified as LSBG by only one author were reinspected together to make a decision. Since visual inspection is time-consuming, we only resorted to it at the last step and tried as much to reduce the number of candidates shortlisted for visual inspection.

To aid in visual inspection, we used two images for every candidate. We generated images enhancing the low surface brightness features using the `APLpy` package (Robitaille & Bressert 2012) and images downloaded from the DESI Legacy Imaging Surveys Sky Viewer (Dey et al. 2019). Furthermore, the g-band Sérsic models from `Galfit` were also used to visually inspect the quality of the model fitting. Each candidate was then categorised into three classes based on the `Galfit` model fit and the images: LSBG, non-LSBG (Artifacts), or misfitted LSBGs. If the model of the galaxy was fitted correctly and the candidate showed LSBG features, it was classified as an LSBG. If the candidate shows LSBG features but does not fit correctly, we classify it as a misfitted LSBG. Finally, if the candidate does not have features of an LSBG, we classify it as an artefact or non-LSBG.

4.3.7 Metrics for comparing models

The metrics used to compare the models are accuracy, AUROC, TPR and FPR which are described in Sect. 2.7. I refer the reader to Sect. 2.7 for a detailed discussion of these metrics.

4.4 Results

4.4.1 Model performance on the testing set

We have created four models of each transformer, namely LSBG DETR and LSBG ViT, with different hyperparameters to generalise our results for both transformers. Each model was implemented as a regression model to predict the probability of an input being an LSBG, and we set 0.5 as the threshold probability for classifying an input as LSBG. Further, we use an ensemble of the four models as the final model for LSBG DETR and LSBG ViT. Table 4.2 describes the architecture, accuracy and AUROC of all the models, including the ensemble models on the test dataset, as mentioned in section 4.2.3.

As mentioned earlier, the more insightful metrics are the true positives (TPR) and the false positives (FRP) rather than overall accuracy. These metrics can be visualised using a confusion matrix, which is shown in Fig. 4.2, for the ensemble models using a threshold of 0.5. The LSBG DETR ensemble had a TPR of 0.96 and an FPR of 0.07, indicating that the LSBG DETR ensemble model can accurately identify 96% of all LSBGs in the DES data, with an estimated 7% contamination rate in the predicted sample. Similarly, the LSBG ViT Ensemble model can identify 97% of all the LSBGs in DES but with 11% contamination.

The receiver operator characteristic (ROC) curve of the LSBG DETR and LSBG ViT ensemble models are shown in Fig. 4.3. In terms of accuracy and AUROC, the LSBG DETR models performed slightly better than the LSBG ViT models. It is clear from Fig. 4.3 that both the ensemble models have a TPR $\sim$ 0.75 even for a high threshold such as 0.9. Indicating that both the ensemble models can confidently identify around $\sim$ 75% of all the LSBGs in DES and assign these candidates with a probability greater than 0.9.

TABLE 4.2: Table comprising the architecture, accuracy, true positive rate (TPR), false positive rate (FPR) and AUROC of all the models in chronological order of creation.

Model name	Accuracy (%)	TPR	FPR	AUROC
LSBG VISION 1	93.55	0.97	0.12	0.980
LSBG VISION 2	93.79	0.97	0.11	0.980
LSBG VISION 3	93.47	0.97	0.11	0.981
LSBG VISION 4	93.51	0.97	0.11	0.980
LSBG VISION Ensemble	**93.75**	**0.97**	**0.11**	**0.983**
LSBG DETR 1	94.36	0.97	0.09	0.982
LSBG DETR 2	94.28	0.96	0.08	0.980
LSBG DETR 3	94.36	0.96	0.08	0.982
LSBG DETR 4	94.24	0.95	0.07	0.982
LSBG DETR Ensemble	**94.60**	**0.96**	**0.07**	**0.984**

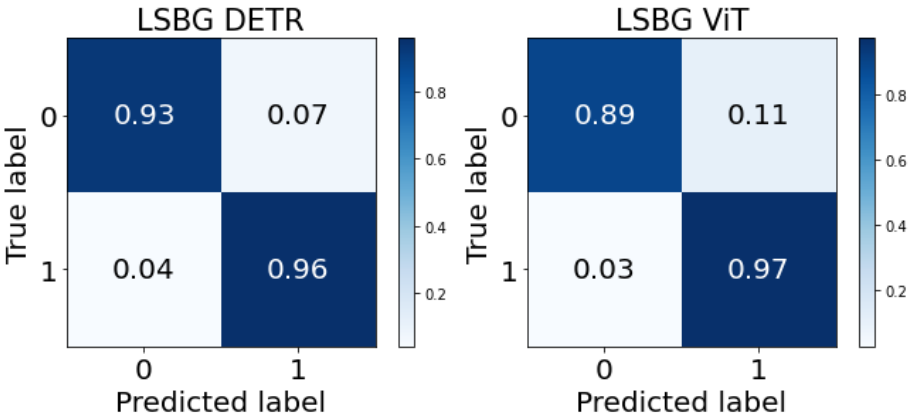


FIGURE 4.2: Confusion matrix of LSBG DETR and LSBG ViT models plotted for a threshold = 0.5. Class 0 represents the artefacts, and Class 1 represents the low surface brightness galaxies.

4.4.2 Search for LSBGs in the full coverage of DES

Since the LSBG DETR model and the LSBG ViT model have different architectures and feature extraction principles, we regard the ensemble models of these two as separate independent transformer classifiers. In order to search for new LSBGs from DES, we employed the transformer ensemble model on the 419 782 objects that satisfied the selection criteria defined in section 4.2.2. The candidates scoring above the threshold probability of 0.5 were catalogued as potential LSBG candidates. The LSBG DETR ensemble classified 27 977 objects as LSBGs, among which 21 005 were already identified by Tanoglidis et al. (2021b). Similarly, the LSBG ViT ensemble classified 30 508 objects as LSBGs, among which had 21 396 LSBGs common with the sample identified by Tanoglidis et al. (2021b). So finally, 6 972 and 9 112 new candidates were classified as potential LSBGs by the LSBG DETR and LSBG ViT ensembles, respectively. However, only the 6 560 candidates identified by both the ensemble models independently were considered for further analysis to reduce the false positives. Since there is a possibility that there might be duplicates of the same candidates existing in the selected sample, we ran an automated spatial crossmatch to remove duplicate objects separated by < 5". The origin of these duplicates can be traced back to the fragmentation of larger galaxies into smaller parts by `SourceExtractor`. After removing the duplicates, the number of potential LSBG candidates reduced from 6 560 to 6 445. As

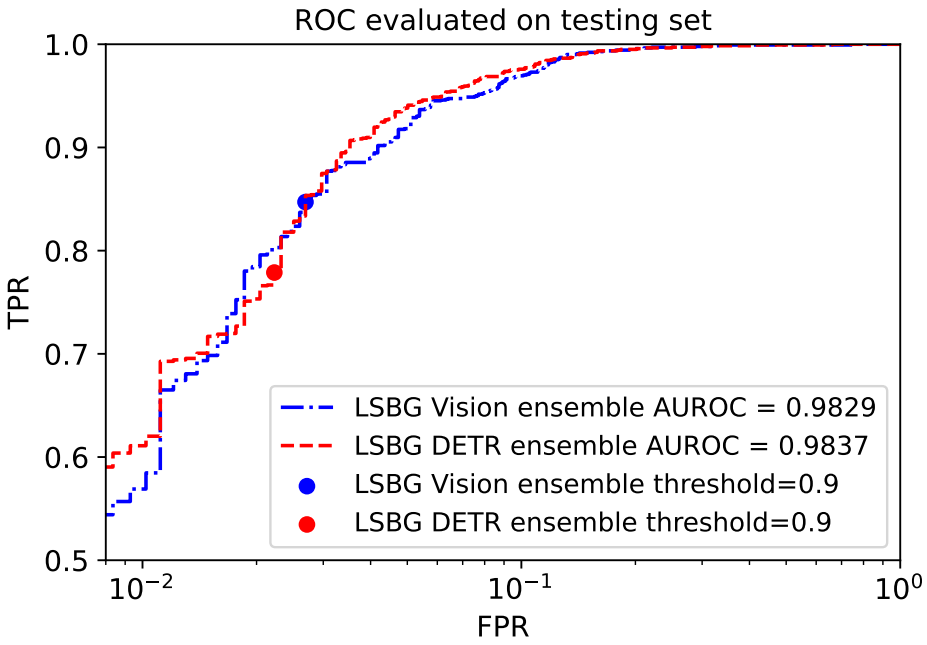


FIGURE 4.3: Receiver operating characteristic (ROC) curve of the ensemble models. The red and blue lines represent the variation of FPR and TPR as a function of the threshold for LSBG DETR and LSBG Vision ensembles, respectively. The red and blue points mark the TPR and FPR for a threshold = 0.9.

discussed in Sect. 4.3.5, these candidates were subjected to single component Sérsic model fitting using `Galfit`.

During the `Galfit` modelling, 999 candidates had failed fits and were consequently removed from the sample since our objective is to produce a high-purity sample with accurate Sérsic parameters. We visually inspected these unsuccessful fits and found that in most cases the presence of a very bright object near the candidate was the cause of the poor Sérsic fit. Of the remaining 5 446 candidates, 4 879 passed the $\bar{\mu}_{eff}$ and $r_{1/2}$ selection criteria outlined in Sect. 4.2.2 with the updated parameters. These 4 879 candidates were inspected visually to identify the genuine LSBGs. After independent visual inspections by the authors, 4 190 candidates were classified as LSBGs and 242 candidates were found to be non-LSBGs. During visual inspection, 447 candidates were found to be possible LSBGs with unreliable measurements from `Galfit`. These candidates are excluded from our final sample, and here we only report on the most confident candidates that were identified as LSBGs during visual inspection. After correcting for the Galactic extinction correction, our final sample was reduced to 4 083 new LSBGs from DES DR1. The schematic diagram showing the sequential selection steps used to find the new LSBG sample is shown in Fig. 4.4. A sample catalogue comprising the properties of the newly identified LSBGs is shown in table 4.3, and some examples of the new LSBGs that we have found are plotted in Fig. 4.5.

The distribution of the $r_{1/2}$, $\bar{\mu}_{eff}$, Sérsic index (n), and axis ratio (q) of the new sample of LSBGs is plotted in Fig. 4.6. The majority of the LSBGs in this new sample have $r_{1/2} < 7''$, and $\bar{\mu}_{eff} < 26$mag arcsec^{-2}. The Sérsic index of the new LSBG sample predominantly lies between 0.5 and 1.5 and has a median value of 0.85. This pattern is similar to the trend identified by Poulain et al. (2021) in the case of dwarf ellipticals, suggesting that a significant portion of the LSBGs sample could be comprised of such sources. In the case of the axis ratio, the new LSBG sample has a median axis ratio of 0.72 and has a distribution lying in the range of 0.3 to 1. The median value of 0.72 suggests that most galaxies in this

sample have a slightly flattened or elongated shape. A detailed discussion of the properties of the new LSBGs identified in this work and their comparison with LSBGs identified by Tanoglidis et al. (2021b) is presented in Sect. 4.5.

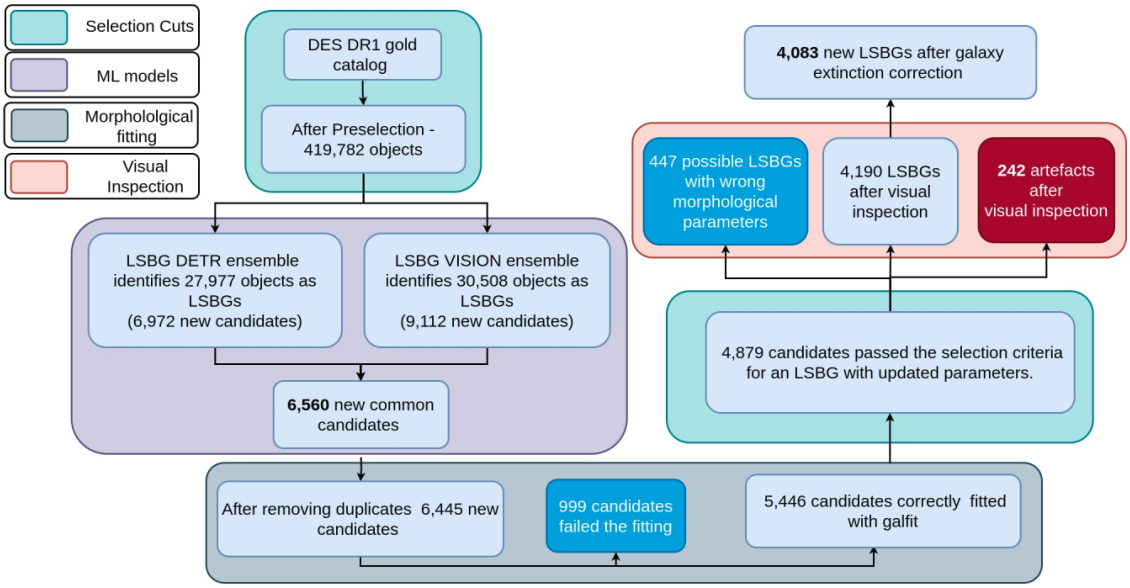


FIGURE 4.4: Schematic diagram showing the sequential selection steps used to find the new LSBG sample.

4.5 Discussion

4.5.1 Transformers as LSBG Detectors

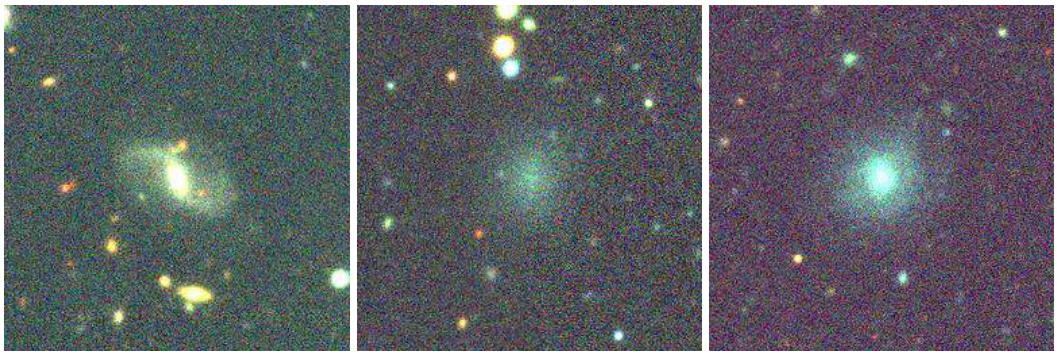
In this study, we introduce the use of transformers as a classifier model for finding the undiscovered LSBGs in DES. Currently, in the literature, one of the reported deep-learning-based models for classifying LSBGs and artefacts is a CNN model named DeepShadows created by Tanoglidis et al. (2021a). They used the catalogue of LSBGs and artefacts identified from DES reported in Tanoglidis et al. (2021b) to generate the training data. The DeepShadows model achieved an accuracy of 92% in classifying LSBGs from artefacts and had a true positive rate of 94% with a threshold of 0.5. Moreover, the DeepShadows model also achieved an AUROC score of 0.974 on this training dataset. However, the DeepShadows was not applied to the complete DES data and checked how it would perform. Nevertheless, DeepShadows was the first deep-learning model to classify LSBGs and artefacts. In addition, Tanoglidis et al. (2021a) also showed that the DeepShaodws was a better classifier than the support vector machine or random forest models. However, in our work, all of our transformer models were able to surpass the DeepShadows model in every metric individually, which can be seen from Table 4.2. Namely, in their respective classes, LSBG DETR 1 and LSBG ViT 2 had the highest accuracies (94.36% and 93.79%), respectively.

Earlier searches for LSBGs have used semi-automated methods such as pipelines based on `imfit` by Greco et al. (2018) or simple machine-learning models such as SVMs by Tanoglidis et al. (2021b). However, the success rate of these methods was very low, and the final candidate sample produced by these methods had around 50% false positives, which had to be removed by visual inspection. Here we explore the possibilities of transformer architectures in separating LSBGs from artefacts. We used two independent ensemble models of LSBG DETR and LSBG ViT models and single component Sérsic model fitting to filter

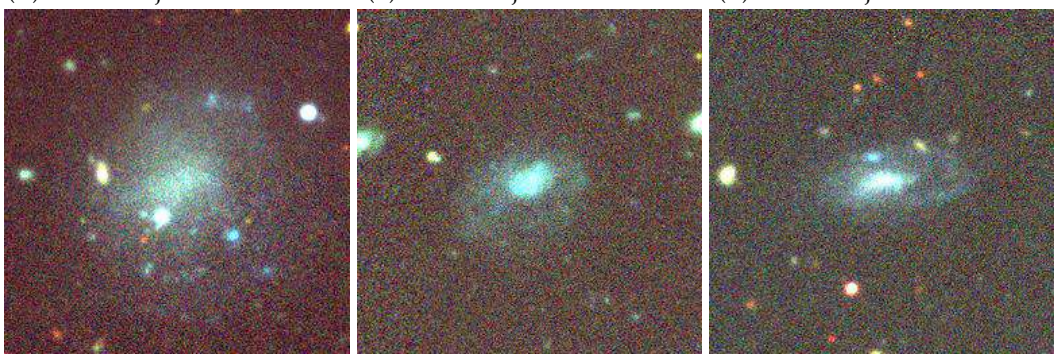
TABLE 4.3: Sample of new LSBGs identified in this work.

COADD_ID	RA (deg)	DEC (deg)	g_{gf}	g_{cor}	$\bar{\mu}_{g\_eff\_gf}$ (mag arcsec^{-2})	$r_{g1/2}$ (arcsec)	n	q	$log_{10}(\Sigma_{star})$ ($M_\odot$ kpc^{-2})	χ^2_{vg}	i_{gf}	i_{cor}	$\bar{\mu}_{i\_eff\_gf}$ (mag arcsec^{-2})	$r_{i,1/2}$ (arcsec)	χ^2_{vi}
61456395	29.7062	-60.4882	19.06	18.97	25.41	9.41	2.17	0.62	5.86	1.06	18.72	18.67	25.37	10.86	1.03
61508029	29.925	4.60483	19.84	19.67	24.33	3.43	0.72	0.85	6.82	1.03	19.27	19.19	23.52	3.06	0.99
61580602	30.3125	-58.1927	18.89	18.83	24.56	5.99	1.5	0.82	6.87	1.02	18.07	18.04	23.72	5.93	1.02
61638403	29.9539	4.75811	20.02	19.87	24.28	3.84	0.74	0.54	6.52	1.04	19.61	19.53	23.62	3.45	1
61638933	29.4862	4.74468	19.2	19.06	24.24	4.59	0.93	0.79	6.82	0.98	18.7	18.63	23.38	3.89	0.99
61712539	30.2824	5.28992	19.67	19.53	24.26	4.57	1.01	0.52	6.58	1.01	19.47	19.4	23.2	3.07	1.03
61766250	30.1112	-9.24769	20.76	20.69	26.29	7.8	1.47	0.42	5.87	1.13	20.19	20.15	25.2	6.15	1.1
62011325	29.9449	-6.78979	21.6	21.51	25.91	3.17	0.5	0.84	6.27	1.02	20.8	20.76	25.21	3.31	1
62053525	29.8595	-16.8586	19.87	19.8	24.24	3.49	0.87	0.73	6.81	1.03	19.32	19.29	23.41	3.06	1.02
62071354	29.8342	-17.1604	20.53	20.46	24.47	3.01	0.84	0.66	6.71	0.99	19.91	19.87	23.65	2.75	0.99
62227622	29.5063	-13.3511	20.84	20.79	24.88	2.68	0.68	0.91	7.24	1.02	19.97	19.94	22.96	1.66	1.05
62371677	29.6675	-5.35506	19.7	19.62	24.25	3.54	0.79	0.84	6.97	1.05	19	18.96	23.32	3.2	1.03
62646182	30.4643	-24.0018	18.46	18.4	24.44	6.88	1.19	0.83	6.77	1	17.98	17.95	23.59	5.8	0.97
62830903	29.9594	-28.8656	20.11	20.07	24.35	2.94	0.71	0.91	7.56	1.03	19.01	18.99	22.44	2.03	1.04
62840965	30.1905	-6.18018	20.18	20.1	24.35	2.74	0.76	0.98	7.35	1.01	19.22	19.18	22.85	2.14	1
63037584	28.9741	-59.5261	19.46	19.39	24.4	6.33	0.9	0.38	6.31	1.01	18.88	18.85	23.99	6.81	1.01
63097874	29.3755	-61.1452	20.6	20.51	24.87	3.08	1.02	0.85	6.58	1.02	20.29	20.25	23.86	2.23	1.02
63113174	29.1535	-61.4326	18.81	18.72	24.78	6.73	0.8	0.86	6.47	1	18.52	18.47	24.15	5.76	0.99
63262376	30.2929	-24.9787	21.48	21.43	26.1	3.92	0.98	0.73	6.18	1.03	20.62	20.6	25.35	4.12	1.04
63527438	30.1134	-4.56234	18.8	18.73	24.51	6.72	0.42	0.68	6.2	1.02	18.68	18.65	24.37	6.63	0.99
63716543	30.0657	-39.5446	20.58	20.52	25.1	3.73	1.16	0.74	6.67	1.02	19.84	19.81	24.01	3.17	1.01
63922768	29.5244	-32.7958	18.41	18.35	24.6	7.62	1.61	0.83	6.59	1.08	18.06	18.03	23.87	6.37	1.04
64480503	29.4619	-23.0107	18.98	18.94	24.33	5.05	0.91	0.86	6.81	1.04	18.5	18.48	23.54	4.39	1
64560481	30.0521	-5.09785	18.18	18.1	24.36	7.21	1.17	0.91	6.86	1.06	17.66	17.62	23.49	6.14	1.01

TABLE 4.4: 'COADD_ID' is the unique id of the source, and 'RA' and 'DEC' gives the sky coordinates of the source as estimated from DES Y3 gold catalogue (Sevilla-Noarbe et al. 2021). Columns 'g_{gf}', 'g_{cor}', '$\bar{\mu}_{g\_eff\_gf}$', and '$r_{g1/2}$' represents the magnitude in g band, the g band magnitude after correcting for Galactic extinction, mean surface brightness and the half-light radius for the g-band fitting using `Galfit`, respectively. The columns 'n','q', and $log_{10}(\Sigma_{star})$ represent the Sèrsic index, axis ratio, and the stellar mass density, respectively. Column 'χ^2_{vg}' represent the reduced chi-square value for the g-band fitting using `Galfit`. Similarly, columns 'i_{gf}', i_{cor}', '$\bar{\mu}_{i\_eff\_gf}$','$r_{i1/2}$', and 'χ^2_{vi}' represents the magnitude in i band, the i band magnitude after correcting for Galactic extinction, the mean surface brightness, the half-light radius and the reduced chi-square value for the i band fitted using `Galfit`, respectively.



(A) Coadd Object Id - 295747204 (B) Coadd Object Id - 62646182 (C) Coadd Object Id - 64560481



(D) Coadd Object Id - 67813078 (E) Coadd Object Id - 69253856 (F) Coadd Object Id - 70739980

FIGURE 4.5: Cutouts of 6 confirmed new LSBGs after visual inspection. The unique identification number (co object id) for each galaxy in DES DR1 is given below each image. The images were generated by combining the g, r and z bands using `APLpy` package (Robitaille & Bressert 2012), and each image corresponds to a 67.32″ × 67.32″ region of the sky with the LSBG at its centre.

the LSBG candidates. Our final sample contained only $\sim$ 5% as non-LSBGs, which is a significant improvement compared to the previous methods in the literature. Following the definition of an LSBG as described in Tanoglidis et al. (2021b), we identified 4 083 new LSBGs from DES DR1, increasing the number of identified LSBGs in DES by 17%. Our results highlight the significant advantage of using deep-learning techniques to search for LSBGs in the upcoming large-scale surveys.

To have more insights into the fraction of false positives from our method, we checked the performance of these models during training. We encountered around 7% and 11% of artefacts from the LSBG DETR ensemble and LSBG ViT sample, respectively, during training on the test dataset. However, using a combination of these models, we reduced the artefact fraction to less than 5% during visual inspection. Most of the non-LSBGs we encountered during the visual inspection were faint compact objects that blended in the diffuse light from nearby bright objects. We use the term 'non-LSBG' instead of artefacts here because, during the visual inspection, we classified some potential LSBGs as non-LSBG; these are objects for which the g-band images had instrumental artefacts or lack of sufficient signal in the g-band. Since the machine learning model intakes three bands as input (g, r and z), this suggests that the model was able to study and generalise the nature of LSBGs in each band and was able to predict if it is an LSBG or not based on the signal from the other bands. However, since we define LSBGs based on their g-band surface brightness and radius in this work, we classified the galaxies without reliable g-band data as non-LSBGs. Some non-LSBGs we encountered during the visual inspection are shown in Fig. 4.7 and Fig. 4.8. With the upcoming surveys of deeper imaging, these

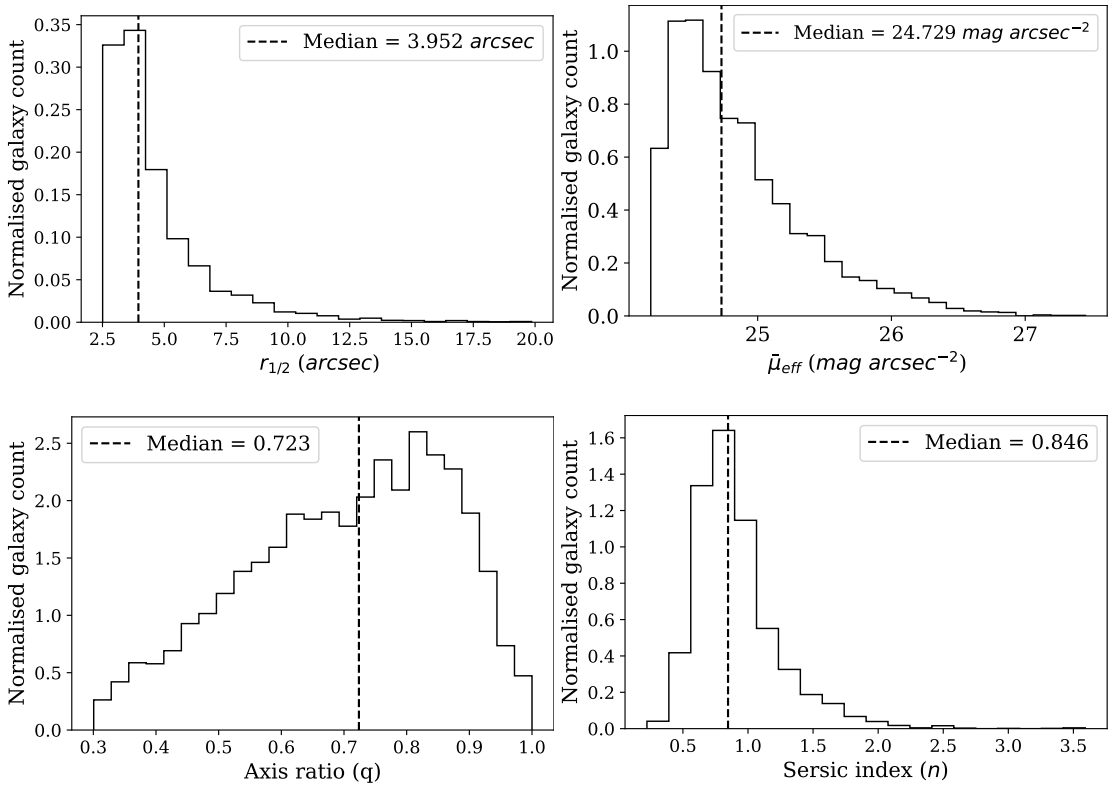


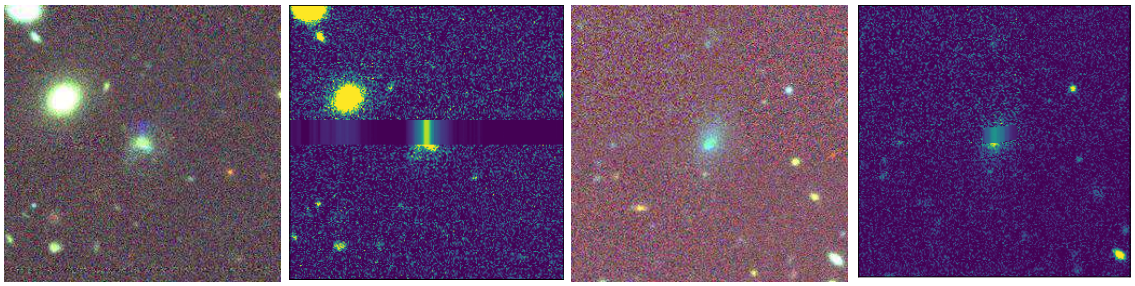
FIGURE 4.6: Normalised distribution of half-light radius (top left panel), mean surface brightness (top right panel), Sérsic index (bottom left panel) and axis ratio (bottom right panel) of the new sample of LSBGs. The dashed line shows the median of the distribution

.

galaxies might be classified as LSBGs which might further reduce the non-LSBGs in our candidate sample.

One another fact to notice when discussing the non-LSBGs from the candidate sample is that some of the candidates identified as LSBGs by the ensemble models (567 out of 5 446) did not meet the selection criteria of being an LSBG after being fitted with `Galfit`. These galaxies had $r_{1/2}$ ranging from 2" to 20", with a median of 3.85", which is similar to the new LSBG sample we found. However, the majority of these galaxies have a mean surface brightness between $24.0 - 24.2$mag arcsec^{-2}, with a median of 24.16mag arcsec^{-2}. This suggests that the machine learning model understood the criteria for angular size for LSBGs during its training, but it did not learn the strict conditions about the surface brightness. This situation is similar to a human expert analysing a galaxy image to determine whether it is an LSBG or not. Features such as the size of the galaxy are easily identifiable to the human eye. However, determining the surface brightness accurately with only the human eye would be challenging, and there may be possible errors near the threshold region, similar to our machine learning model. So one could say that the machine learning model is behaving approximately like a human visual expert.

Judging from the performance of our model on the training data, we cannot assert that we have discovered all the possible existing LSBGs from the DES DR1. As we can see from Fig. 4.2, the TP rate for the individual ensemble models were 0.96 and 0.97, respectively. This means that the model has not found all the possible LSBGs and a minor fraction of LSBGs is yet to be found in DES DR1. Moreover, to reduce the FPR and the burden during the visual inspection, we have only visually inspected the candidates identified commonly



(A) (B) (C) (D)

FIGURE 4.7: Examples of candidates (Coadd object id - 149796289 and 374192591) classified as non-LSBGs during visual inspection because of glitches in the g-band near the galaxy. The panels (a) and (c) show the RGB image created using the g, r and z bands with `APLpy` package (Robitaille & Bressert 2012). The panels (b) and (d) show the image in the g band. Each image corresponds to a 67.32″ × 67.32″ region of the sky with the candidate at its centre.

(A) Coadd object id - 251235955 (left) and 99585243 (right)

(B) Coadd object id - 125313682 (left) and 113818243 (right)

FIGURE 4.8: Examples of candidates classified as non-LSBG during visual inspection because of lack of sufficient signal in the g-band (a) are shown in the top panel. Candidates classified as non-LSBG during visual inspection because of being artefacts are shown in the lower panel (b). The RGB images are created using the g, r and z bands with `APLpy` package (Robitaille & Bressert 2012). Each image corresponds to a 67.32″ × 67.32″ region of the sky with the candidate at its centre.

by both the ensemble models and passed the criteria for correctly fitting by `Galfit`.

One of the notable things about the models in this work is that we are using two different ensemble models, each having four models in the ensemble. As we mentioned earlier, each ML model can be considered equivalent to a human inspector, and the ensemble models help balance out the disadvantages of the other models in the ensemble. A closer look at the individual probability distribution of these modes shows that there are 310 candidates among the 4 083 confirmed LSBG candidates, which had a probability of less than 0.5 for at least one model among the individual models. However, since we used an average ensemble model, we were able to identify these LSBG by balancing out the probability, which shows the advantages of using an ensemble model over a single model.

Here, we use the visual inspection as the final step to confirm the authenticity of an LSBG detected by the models. However, it is essential to acknowledge the potential for human bias during the visual inspection, which can impact the accuracy and reliability of the results. For example, during the visual inspection, the visual inspectors disagreed on labelling approximately 10% of the candidate sample. Most of these galaxies had a

mean surface brightness greater than 25.0mag arcsec^{-2}, which suggests that even for human experts, it is challenging to characterise extremely faint LSBGs. However, with better imaging, this might change, but we must acknowledge that there will always be some human bias and error associated with human inspection. Also, we must consider that in the upcoming surveys, such as LSST or Euclid, the data size will be too large to inspect visually. In this scenario, relying solely on improved automated methods to purify the sample and accepting a small fraction of false positives could be a feasible solution.

4.6 The new sample of DES DR1 LSBGs

4.6.1 The newly identified LSBG sample

The optical color of a galaxy can give insights into its stellar population. Conventionally, based on their color, the galaxies are divided into red and blue galaxies, and it has been known that color is strongly correlated to the morphology of a galaxy (Strateva et al. 2001). Blue color galaxies are usually found to be highly active star-forming spiral or irregular systems. In contrast, red color galaxies are mostly found to be spheroidal or elliptical. In addition, the red galaxies have also been found to tend to cluster together compared to the blue galaxies (Bamford et al. 2009).

The LSBGs found by Tanoglidis et al. (2021b) have found a clear bimodality in the $g-i$ color distribution similar to the LSBGs found by Greco et al. (2018). The $g-i$ color distribution of the 4 083 new LSBGs and the 23,790 LSBGs found by Tanoglidis et al. (2021b) is shown in In Fig. 4.9. We follow the criteria defined by Tanoglidis et al. (2021b) to define red galaxies as galaxies having $g-i>0.6$ and blue galaxies as galaxies having $g-i<0.6$ where g and i represent the magnitude in each band. The color as a function of mean surface brightness in g-band for the new sample identified in this work and the LSBGs identified by Tanoglidis et al. (2021b) is shown in Fig. 4.10. There are 1112 red LSBGs and 2,944 blue LSBGs in the new LSBG sample. [4] From Fig, 4.9, we can see that we have identified a relatively large fraction of blue LSBGs compared to Tanoglidis et al. (2021b) and a considerable fraction of new red LSBGs with $g-i\geq 0.80$ and a mean surface brightness less than 25.0mag arcsec^{-2}. The bias against blue LSBGs and highly red LSBGs in the sample created by Tanoglidis et al. (2021b) may have been caused by the bias in the training set used to create the SVM, which preselected the LSBG candidates. This bias could have occurred because a large fraction of their training set consisted of LSBGs near the Fornax cluster, which are mainly red LSBGs.

Looking at the distribution of $\bar{\mu}_{eff}$ values of the new sample, both the red and blue LSBGs have a similar mean surface brightness range, with a median $\bar{\mu}_{eff}$ of 24.75 and 24.68mag arcsec^{-2}, respectively. Both populations of red LSBGs and blue LSBGs from the new sample have sizes ranging from 2.5" - 20". However, as mentioned earlier, most of these LSBGs have radii less than 7", with a median of 4.01" for blue LSBGs and 3.59" for red LSBGs. In comparison, blue LSBGs tend to have larger angular radii compared to red LSBGs. The Sérsic index distribution of the red and blue LSBGs in the new sample has similar distribution and almost equal median values (0.847 and 0.845 for red and blue LSBGs, respectively). A median Sérsic index of around 0.84 indicates that the majority of the galaxies are closer to a disk-shaped geometry, irrespective of their color. The distribution of the axis ratio of the red and blue LSBGs from the new sample shows a clear difference, as shown in Fig. 4.11. The median of the axis ratio distribution of the blue and red LSBGs is 0.7 and 0.8, respectively. This indicates that, in general, the red LSBGs are rounder than the blue LSBGs.

[4] 27 LSBGs failed the modelling using `Galfit` for i-band, and they are not included in this color analysis.

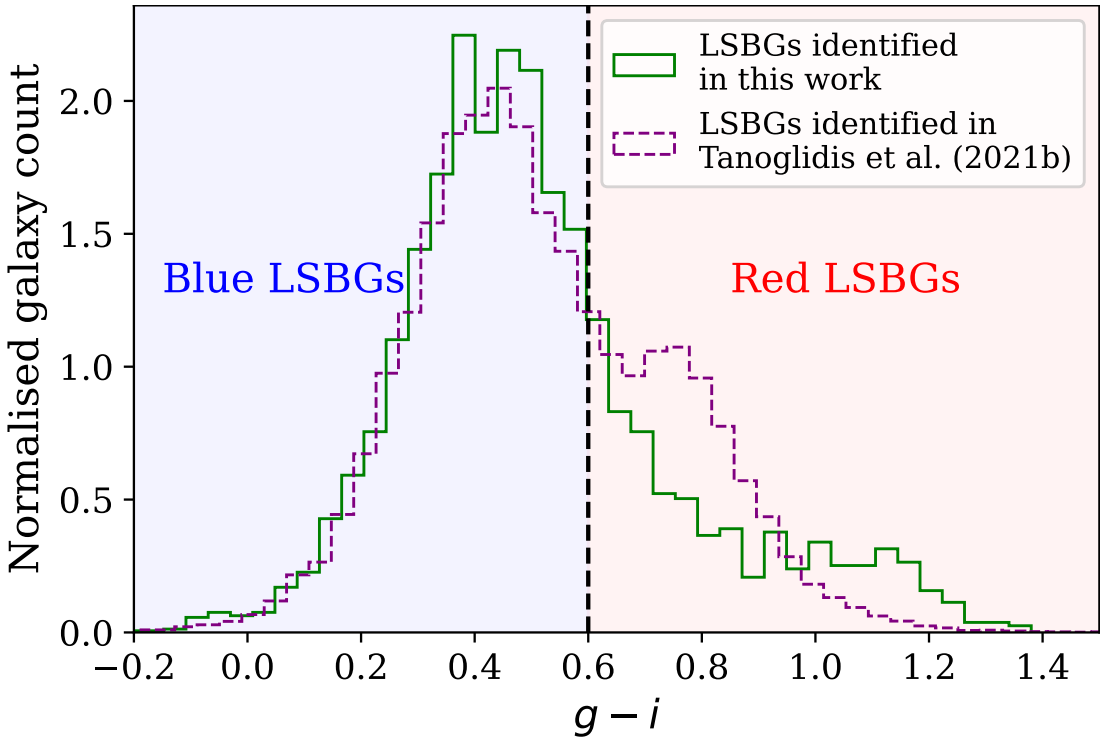


FIGURE 4.9: Normalised distribution of color of the LSBGs from the new sample identified in this work and the LSBGs identified by Tanoglidis et al. (2021b).

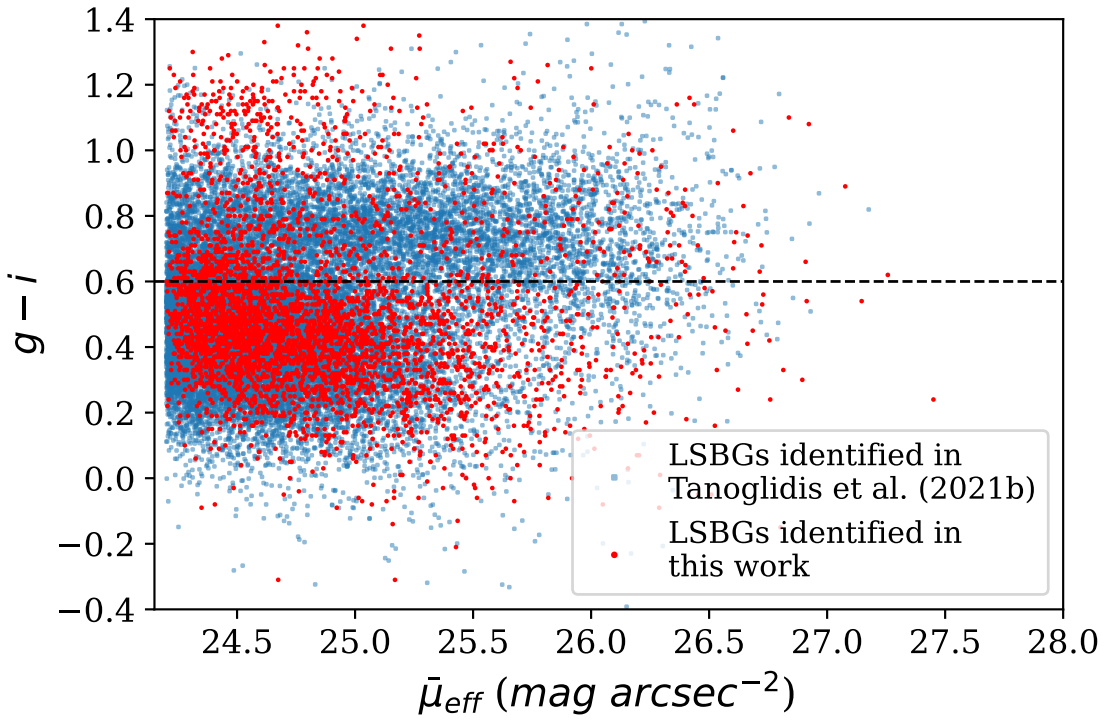


FIGURE 4.10: The color as a function of mean surface brightness in the g-band for the new sample identified in this work and the LSBGs identified by Tanoglidis et al. (2021b). The dashed line shows the separation between red and blue LSBGS.

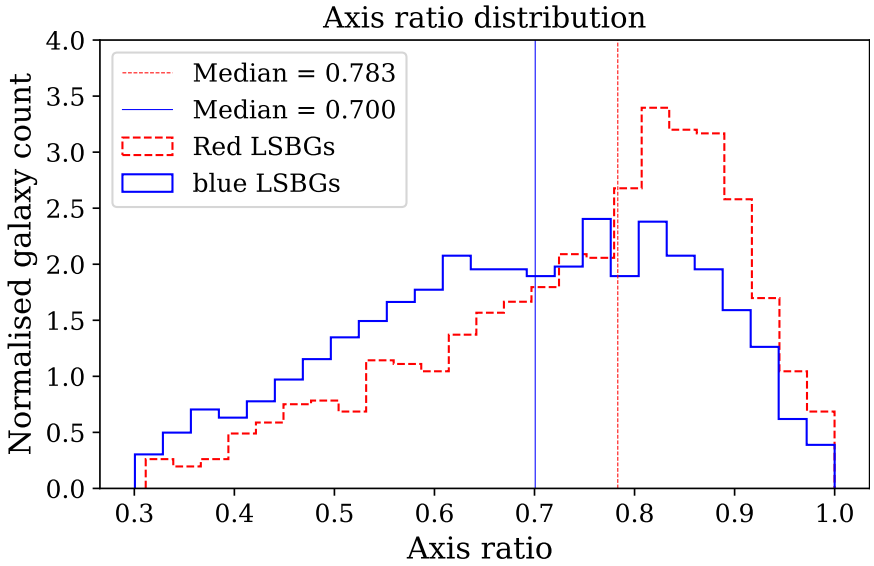


FIGURE 4.11: Normalised distribution of axis ratio (left panel) of red and blue LSBGs from the new sample. The vertical lines show the median for each class.

4.6.2 Why are there additional LSBGs?

One of the other things to investigate at this moment will be how different the new LSBG sample is compared to the LSBGs identified by Tanoglidis et al. (2021b). Or, more precisely, one could wonder why these many LSBGs have been missed previously and whether it is somehow related to the nature of these galaxies. Apart from the Sérsic index, all other Sérsic parameters of the new and the old sample have an almost similar distribution. The distribution of the Sérsic index for the new sample identified in this work and the LSBG sample identified by Tanoglidis et al. (2021b) is shown in Fig. 4.12. The new LSBG sample has a Sérsic index predominantly in the range $n < 1$, which is comparatively lower than the Sérsic index distribution of LSBGs identified by Tanoglidis et al. (2021b). However, this does not point to any reason why these LSBGs were missed in the previous search, and moreover, Tanoglidis et al. (2021b) have also commented on the under-representation of red LSBGs with small Sérsic index in their sample.

To answer the aforementioned question, a close inspection of the methodology of Tanoglidis et al. (2021b) shows that most of the new LSBGs (82%) we identified here were missed by the SVM in their first pre-selection step. This shows the importance of methodology in preselecting the samples. Since the methodology used by Tanoglidis et al. (2021b) and Greco et al. (2018) have considerable similarities (e.g., usage of SVM), this indicates that Greco et al. (2018) might have also missed some LSBGs from the HSC-SSP survey and the fraction should be greater in comparison to Tanoglidis et al. (2021b). It should be noted that there is a slight overlap in the regions of observation by Greco et al. (2018) and DES, as shown in Fig. 4.13. There are 198 LSBGs identified by Greco et al. (2018) from HSC-SSP in the field of view of DES and detected in the DES Y3 gold catalogue. Among these 198 LSBGs, Tanoglidis et al. (2021b) has recovered 183 LSBGs, and we recovered 10 more additional LSBGs from this field, making the total number of recovered LSBGs to 193. We would also like to point out that there are additional LSBGs ($\sim$ 200) in our total sample in the same region, but missed by Greco et al. (2018), even though the HSC-SSP data used by Greco et al. (2018) is about 2 orders of magnitude deeper than the DES DR1. However, we have also missed some LSBGs ($\sim$ 150) that have been identified by Greco et al. (2018). These LSBGs were not detected in the DES Y3 gold catalogue and subsequently

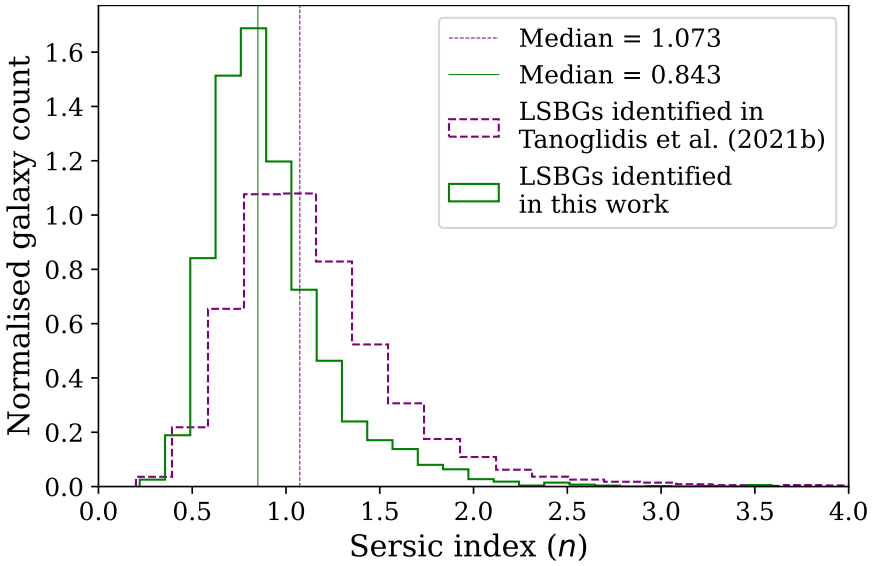


FIGURE 4.12: Normalised distribution of the Sérsic index of the LSBGs identified in this work and by Tanoglidis et al. (2021b). The vertical lines show the median for each class

.

were missed by the searches by Tanoglidis et al. (2021b) and ours. With the DES data release 2 (DES DR 2) having an improved depth ($\sim$ 0.5 mag; Abbott et al. 2021), we should expect an increase in the number of LSBGs from DES. Therefore, there is a potential for using transfer learning with transformers in the future search for LSBGs from DES DR 2 (Abbott et al. 2021) and HSC-SSP data release 3 (Aihara et al. 2022).

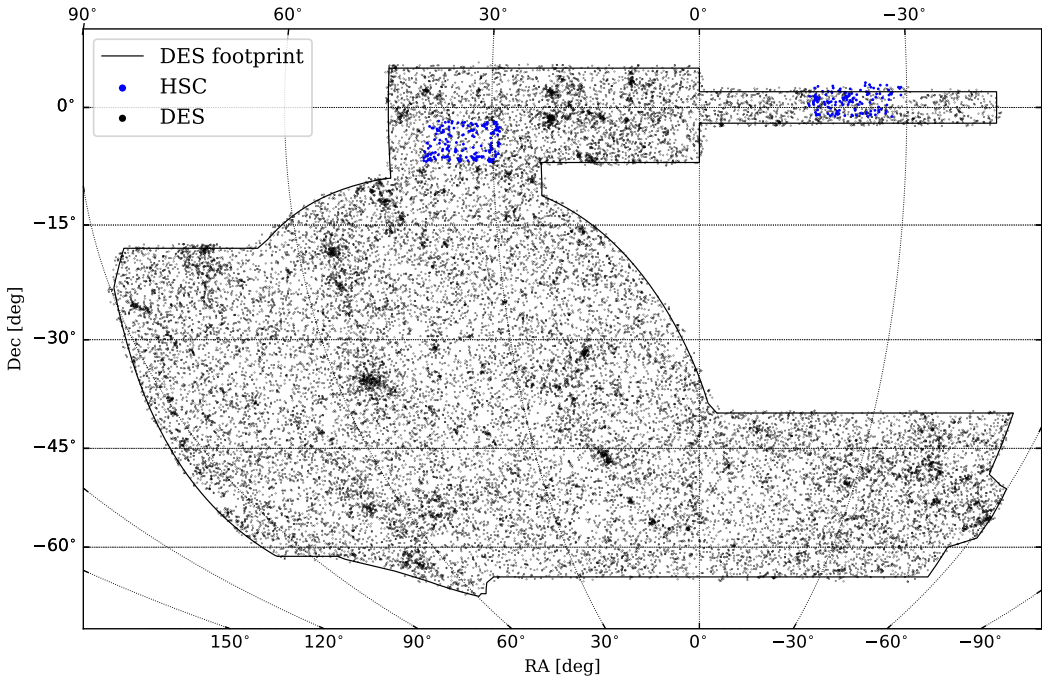


FIGURE 4.13: Sky distribution of the LSBGs identified from DES (black dots) by Tanoglidis et al. (2021b) + this work and the LSBGs identified from HSC-SSP (blue dots) by Greco et al. (2018).

With the addition of the new 4 083 LSBGs, the number of LSBGs in the DES increased to 27,873, effectively increasing the average number density of LSBGs in DES to ~ 5.5 deg^{-2}. In addition, it should also be noted that there are still around $\sim$ 3000 candidates identified by the ensemble models, which have not undergone further analysis to be verified as LSBGs. Potentially indicating that the number of LSBGs in DES might increase further in future. Hence the average number density of 5.5 deg^{-2} reported here can only be taken as a lower limit. Earlier, Greco et al. (2018) estimated that the average number density of LSBGs in HSC-SSP is $\sim$ 3.9 deg^{-2}. However, this estimate was based on LSBG samples with $\bar{\mu}_{eff} > 24.3$mag arcsec^{-2}, unlike the $\bar{\mu}_{eff} > 24.2$mag arcsec^{-2} selection we adopted in this work. For a similar selection on mue > 24.2mag arcsec^{-2} in the combined sample presented here (LSBGs identified in this work + LSBGs identified by Tanoglidis et al. (2021b)), we obtain a higher number density of 4.9 deg^{-2}, compared to the previous estimates (3.9 deg^{-2} from Greco et al. (2018) and 4.5 deg^{-2} from Tanoglidis et al. (2021b)).

As discussed above, the number density of the LSBGs will be influenced by the methodology used to search for the LSBGs. Similarly, one other intrinsic factor that can influence the number density is the completeness of the survey. Improved imaging techniques can reveal fainter objects, leading to an increase in the number density. The completeness of a survey can be determined by plotting the galaxy number count, and one could also have a rough idea about the redshift distribution of the objects of interest by comparing it with the Euclidean number count. Fig. 4.14 shows the number count of LSBGs identified in DES (this work and Tanoglidis et al. (2021b)) and HSC (Greco et al. 2018). As expected, HSC has better completeness than DES. However, HSC still has a smaller number density than DES, which is evident from comparing the peaks of both number counts. The slope of the number counts near 0.6 (representing Euclidean geometry) for both HSC and DES suggest that most identified LSBGs are local (Yasuda et al. 2001). Furthermore, E Greene et al. (2022) has analysed the LSBG sample from HSC and estimated that the 781 LSBGs identified by Greco et al. (2018) have a redshift less than 0.15.

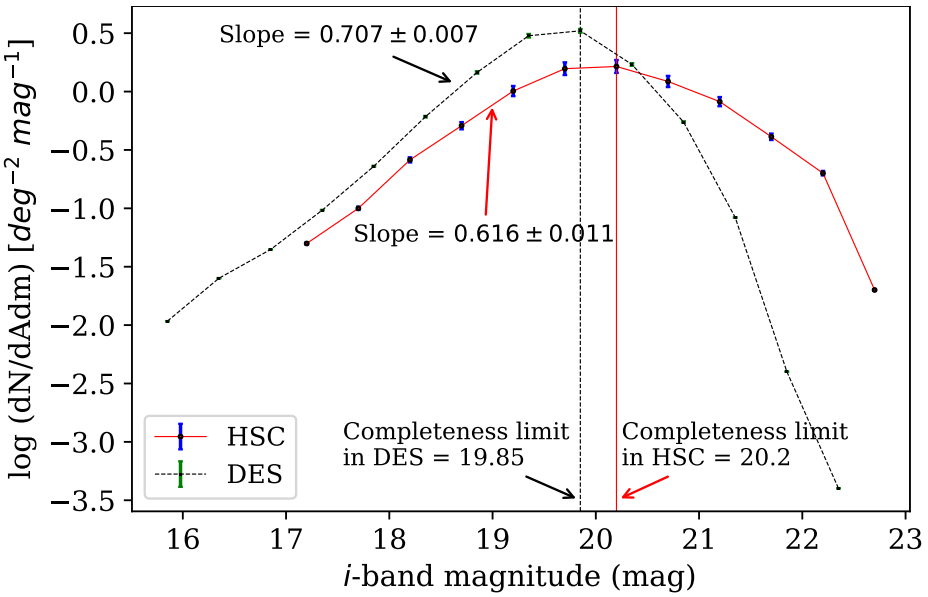


FIGURE 4.14: The number count of galaxies as a function of *i*-band magnitude, with the y-axis displaying the logarithm of the number density per apparent magnitude. The red line with the blue error bars represents the data from HSC, and the black dashed line with green error bars represents the data from DES.

With the increasing number of LSBGs identified from different surveys, one of the

other questions that need to be addressed at this moment is the definition of an LSBG itself. One could define a different definition for an LSBG, consequently leading to finding a completely different sample of LSBGs from the same dataset, which in turn can affect the conclusions of the study. One of the primary factors contributing to these discrepancies is the current reliance on surface brightness-based definitions for LSBGs, which are contingent upon the observation band in use. Different observation bands may involve distinct threshold values. Depending on the band we use, the LSBG definition will likely vary. In this scenario, one potential solution is to define an LSBG based on the stellar mass density of the galaxy. Current definitions based on the stellar mass density define an LSBG as a galaxy with a stellar mass density, $\Sigma_{star} \lesssim 10^7$ M$_\odot$ kpc^{-2} (e.g., Carleton et al. 2023). Following Eq. 1 of Chamba et al. (2022), we made an estimate of the stellar mass surface density using our observed i-band surface brightness $\bar{\mu}_{eff}$ and the stellar mass-to-light ratio obtained from the $g-i$ color (Du et al. 2020). The stellar mass surface density distribution of the LSBGs from DES and HSC-SSP is shown in Fig. 4.15. Here we can see that most of the LSBGs satisfy this condition, and only a small percentage stay above the threshold of 10^7 M$_\odot$ kpc^{-2}. On average, the LSBGs from DES have a higher stellar mass surface density than those from HSC-SSP, which could be attributed to the higher depth in the data used by Greco et al. (2018). However, as argued by Chamba et al. (2022), accurate estimation of the stellar mass density requires deep photometry in multiple bands. In our case, we employed a single color, and as a result, the constraints we derived on the stellar mass density may be limited in accuracy.

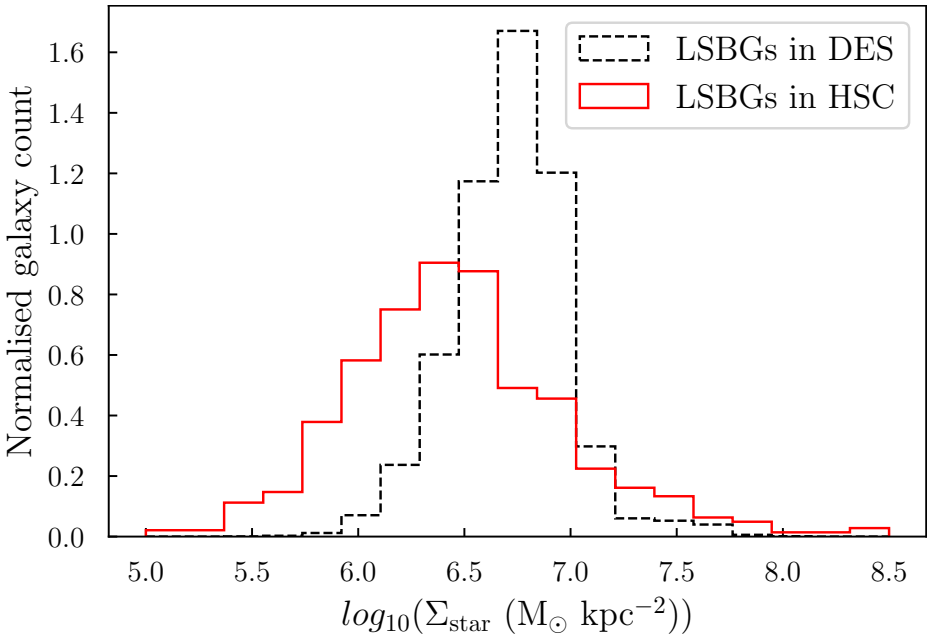


FIGURE 4.15: Normalised distribution of stellar mass surface density of LSBGs identified in HSC (red line) and DES (black line).

4.7 Clustering of LSBGs in DES

The on-sky distribution of the red and blue LSBGs identified in this work, along with those identified by Tanoglidis et al. (2021b), is shown in Fig. 4.16 and Fig. 4.17. In the local universe, "normal" high surface brightness red galaxies tend to cluster together, while blue galaxies are much more dispersed in the field (Zehavi et al. 2005). Such a trend is also

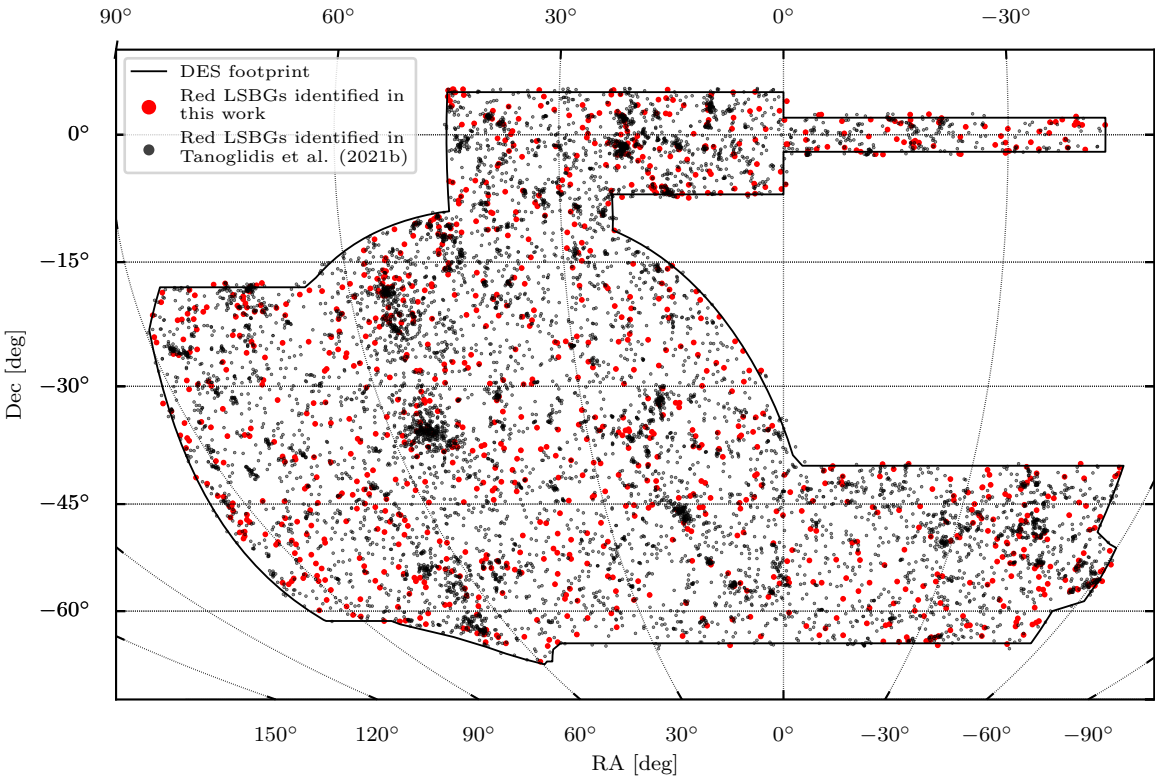


FIGURE 4.16: Sky distribution of red LSBGs identified in this work (red dots) and the LSBGs identified (black dots) by Tanoglidis et al. (2021b).

clearly visible for the LSBG sample. As seen in Fig. 4.16, red LSBGs tend to form concentrated nodes. In contrast, the blue LSBGs are distributed much more homogeneously in the sky, as seen in Fig. 4.17.

A two-point auto-correlation function is a statistical tool commonly used to quantify the galaxy clustering (Peebles 1980). Here we use the angular two-point auto-correlation function, $\omega(\theta)$, computed using the Landy & Szalay (1993) estimator defined as

$$\omega = \frac{\hat{DD}(\theta) - 2\hat{DR}(\theta) + \hat{RR}(\theta)}{\hat{RR}(\theta)}, \tag{4.5}$$

where

$$\hat{DD} = \frac{DD(\theta)}{n_d(n_d-1)/2}, \tag{4.6}$$

$$\hat{DR} = \frac{DR(\theta)}{n_d n_r}, \tag{4.7}$$

$$\hat{RR} = \frac{DD(\theta)}{n_r(n_r-1)/2}. \tag{4.8}$$

Here DD(θ) is the number of pairs in the real sample with angular separation θ, RR(θ) is the number of pairs within a random sample, DR(θ) is the number of cross pairs between the real and random samples, n_d is the total number of real data points, and n_r is the total number of random points.

We use a random sample of 4 491 746 points generated from the DES footprint mask. To compute $\omega(\theta)$ we employ `treecorr` (Jarvis 2015). Errors are estimated using jackknife resampling where the sky is divided into 100 equal-sized batches for resampling (Efron &

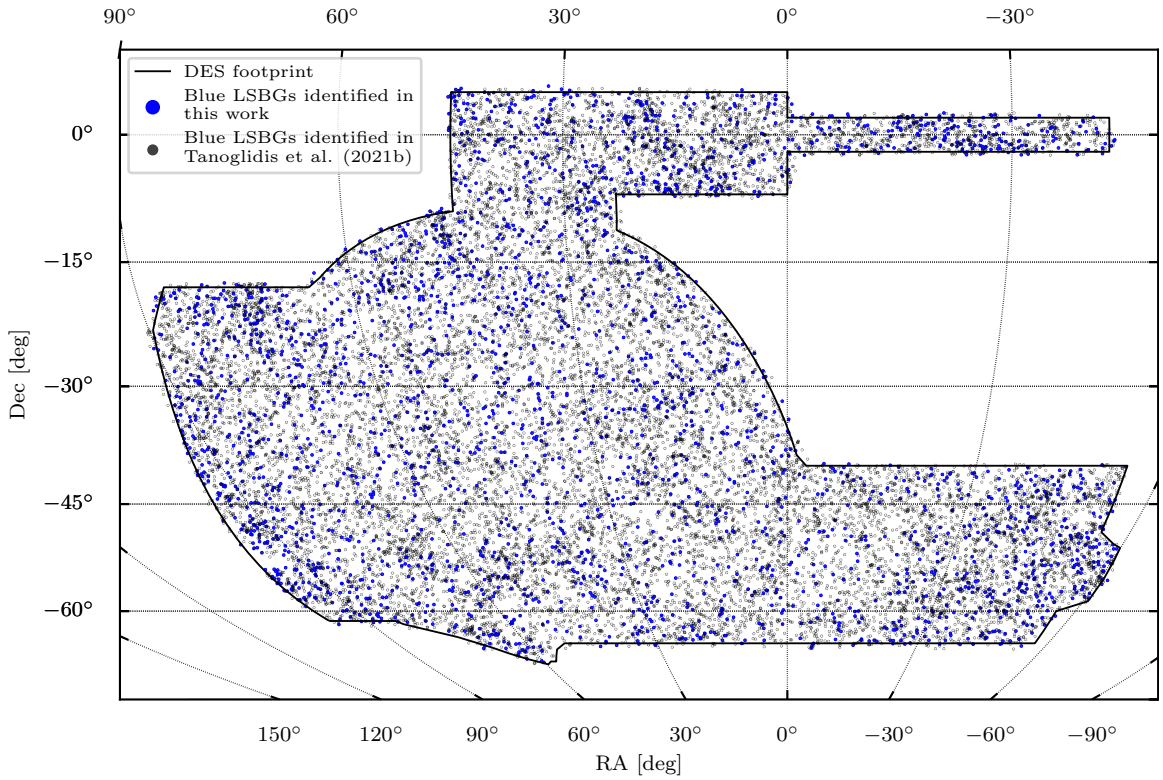


FIGURE 4.17: Sky distribution of blue LSBGs identified from the new sample (blue dots) and the LSBGs identified (black dots) by Tanoglidis et al. (2021b).

Gong 1983). For high surface brightness galaxy samples, the angular correlation function very often can be well fitted by a single power-law (Peebles & Hauser 1974; Peebles 1980; Hewett 1982; Koo & Szalay 1984; Neuschaefer et al. 1991)

$$\omega(\theta) = A\theta^{1-\gamma} \tag{4.9}$$

where A is the amplitude which represents the strength of the clustering, and γ represents the rate at which the strength of the clustering reduces as we go to large angular scales. This power-law behaviour is usually observed in a wide range of angular scales; however, it is not universal, especially on the smallest scales. Full modelling of the shape of the correlation function requires taking into account different processes governing galaxy clustering on small scales (corresponding to galaxies located in the same dark matter halo) and at larger scales (corresponding to clustering of different haloes). This modelling is usually done using the halo occupation distribution models (HOD) (Ma & Fry 2000; Peacock & Smith 2000; Zheng et al. 2005; Kobayashi et al. 2022). In this work, however, we perform only a preliminary analysis and base interpretation of our data on the power-law fitting only.

To compare the clustering of the LSBGs with the clustering of the high surface brightness galaxies (HSBGs), we constructed a control sample of HSBGs from the DES data. For this purpose, we selected galaxies in the surface brightness range $20.0 < \bar{\mu}_{eff} < 23$mag arcsec^{-2} and in the magnitude range $17 < g < 23$ mag (which is the same magnitude range as our LSBG sample). Additionally, we applied a photometric redshift $z < 0.1$ cut in order to keep the HSBGs sample consistent with the LSBGs, which are also expected to be mostly local (E Greene et al. 2022). For this purpose, we used the photometric redshifts from the DES Y3 gold catalogue calculated using the Directional Neighbourhood Fitting (DNF) algorithm (Sevilla-Noarbe et al. 2021; De Vicente et al. 2016). In

addition, we also applied the selection cuts on the parameters from `SourceExtractor` such as `SPREAD_MODEL,EXTENDED_CLASS_COADD` and on colors (using the `MAG_AUTO` magnitudes) as described in Sect. 4.2.2.

Initially, we computed the angular two-point auto-correlation function for the samples of LSBGs and HSBGs. Then we split the samples into red and blue galaxies to measure their clustering properties separately. For LSBGs, we followed the criterion defined in Sect. 4.6, i.e. a color cut of $g - i = 0.6$ mag to separate blue and red sources. As seen from the color histogram presented in Fig. 4.18, the HSBGs show a bi-modality around $g - i = 1.0$ mag, which can be most likely attributed to their different stellar masses. Consequently, we use the boundary $g - i = 1.0$ mag to divide our HSBG sample into red and blue subsamples. The properties of all the samples used for the measurement of the galaxy clustering, together with the best-fit power-law parameters, are listed in Table 4.5. The 2-point auto-correlation functions for all the samples described above are shown in Fig. 4.19.

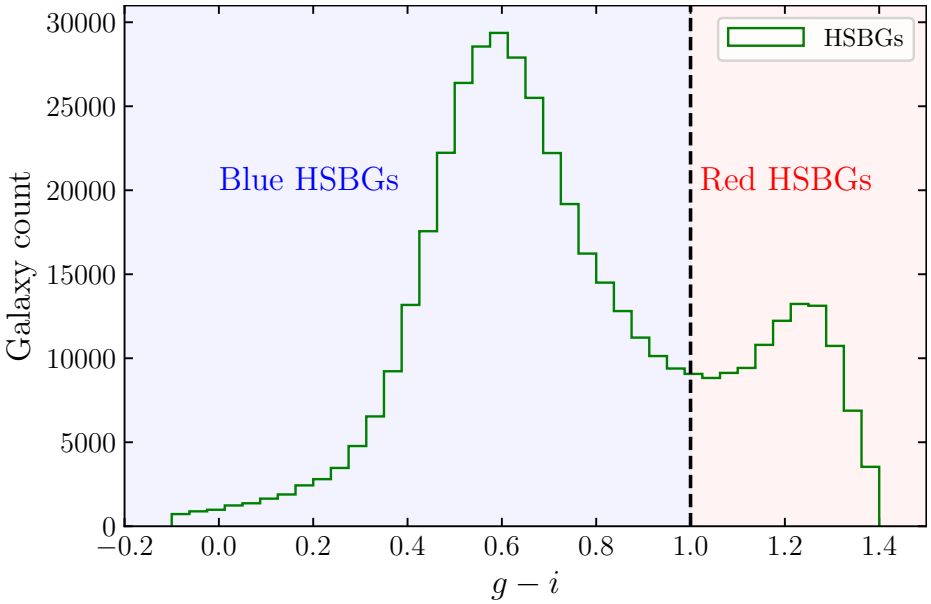


FIGURE 4.18: Color distribution of the HSBGs from the DES DR1. The vertical line at $g - i = 1.0$ shows the color separation of the HSBGs into red and blue galaxies.

As it is clear from Fig. 4.19, the angular two-point auto-correlation function of the red LSBGs does not follow a power law at small angular scales. Therefore, the power-law fits were only performed in the range of 0.15 deg to 7 deg to avoid them being affected by the one-halo effects. In part well fitted by the power law, for the red LSBGs, $\omega(\theta)$ is significantly steeper than for the blue LSBGs. However, it flattens at smaller scales, i.e. between 0.01 deg and 0.3 deg; this behaviour is also transmitted to the full sample of LSBGs. In contrast, the blue LSBGs follow a power law behaviour, with a lower clustering amplitude and a much less steep slope, in almost all the angular scales. This behaviour of the angular correlation function might be explained by the observations by van der Burg et al. (2016) and Wittmann et al. (2017) that the number of LSBGs close to the cores of galaxy clusters decreases. Such suppression may reduce the clustering power on small scales, leading to a flattening in the auto-correlation function, which is seen for the red LSBGs, which are mostly associated with clusters.

Comparison of clustering of the LSBGs and the HSBGs also shows notable differences. Not surprisingly, red samples, both of HSBGs and LSBGs, are more clustered than their

TABLE 4.5: Best-fitting power law parameters for the angular two-point auto-correlation function for HSBG and LSBGs along with the information on the number of galaxies, median g-band magnitude, and the mean surface brightness for each sample.

Sample	Number of galaxies	Median g (mag)	Median $\bar{\mu}_{eff}$ (mag arcsec^{-2})	A	γ
All HSBGs	451,310	18.84	21.66	0.091 $\pm$0.004	1.651 $\pm$0.021
Red HSBGs	103,900	17.96	21.21	**0.245 $\pm$0.004**	**1.848 $\pm$0.012**
Blue HSBGs	347,410	19.21	21.81	0.0648 $\pm$0.004	1.631 $\pm$0.036
All LSBGs	27,840	20.11	24.66	0.138 $\pm$0.013	1.941 $\pm$0.048
Red LSBGs	18,924	20.23	24.89	**0.671 $\pm$0.079**	**2.090 $\pm$0.071**
Blue LSBGs	8,916	20.07	24.59	0.051 $\pm$0.001	1.620 $\pm$0.025

blue counterparts. At the same time, the red LSBG sample has a significantly higher clustering amplitude than the reference red HSBG sample. Red LSBGs also display a steeper slope of $\omega(\theta)$ at angular scales larger than 0.15 deg, but at smaller scales, their $\omega(\theta)$ flattens, unlike in the case of red HSBGs for which we can even observe a hint of an upturn which can be associated with a one-halo term. This picture is consistent with a scenario in which red LSBGs are mostly associated with dense structures like clusters but do not populate their centres but rather the outskirts. In contrast, red HSBGs display the usual behaviour of red passive galaxies, appearing in a variety of environments, with a tendency to cluster and gather most strongly in the cluster centres.

Blue LSBGs have a significantly lower clustering amplitude than their HSBG counterparts. At the same time, the slope of their $\omega(\theta)$ at scales larger than 0.15 deg remains very similar. The blue HSBGs and LSBGs follow the usual distribution of blue star-forming galaxies, dispersed in the field and avoiding clusters. These results are consistent with the results obtained by Tanoglidis et al. (2021b) for their sample of DES LSBGs. They compared the clustering of LSBGs with very bright galaxies in the magnitude range of $14 < g < 18.5$ mag from the 2MPZ catalogue (Bilicki et al. 2014). They found that LSBGs had higher clustering amplitude in the range of 0.1 to 2 degrees, which is similar to our observations.

However, our results contradict the early estimates from Bothun et al. (1993) and Mo et al. (1994), who infer that the LSBGs tend to cluster weakly spatially. However, their analyses were limited by a small data sample ($\sim$400 LSBGs), a small area of the sky, and most likely selection biases. Given the low accuracy of photometric redshifts for LSBGs in our sample, we do not attempt to reconstruct their spatial clustering in this work. Further analysis is planned as a follow-up to this study.

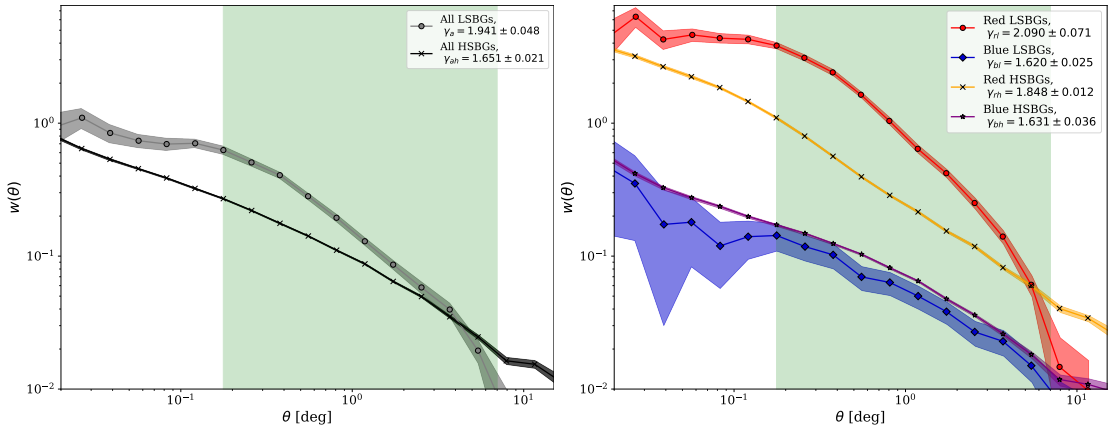


FIGURE 4.19: Angular autocorrelation function for the full sample of LSBGs (grey line with open circles) and the sample of HSBGs (black line with crosses) is shown in the left panel. The angular autocorrelation function of the red LSBGs (red line), blue LSBGs (blue line), red HSBGs (orange line) and blue HSBGs (purple line) is shown in the right panel. The vertical green shaded region represents the region fitted for a power law ($\omega = A\theta^{1-\gamma}$), and the corresponding γ values are shown in the legend.

4.8 Identification of ultra-diffuse galaxies

As discussed in Sect. 4.1, UDGs are a subclass of LSBGs that have extended half-light radii $r_{1/2} \geq 1.5$ kpc and a central surface brightness $\mu_0 > 24$mag arcsec^{-2} in g-band (van Dokkum et al. 2015a). Significant population of UDGs have been discovered in the Coma cluster by van Dokkum et al. (2015a) and other investigations have revealed a large number of UDGs in other galaxy clusters (Koda et al. 2015; Mihos et al. 2015; Lim et al. 2020; La Marca et al. 2022b,a). Later on, studies have shown that thousands of UDGs can be found in single individual clusters and that the abundance of UDGs scales close to linearly with host halo mass (van der Burg et al. 2016; Mancera Piña et al. 2018).

In order to investigate if there are any cluster UDGs in the sample of LSBGs we identified in DES, we crossmatched our total LSBG sample (23,790 LSBGs from Tanoglidis et al. 2021b and the 4 083 new LSBGs we identified) with the X-ray-selected galaxy cluster catalogue from the ROSAT All-Sky Survey (RXGCC; Xu et al. 2022). All the LSBGs at the angular distance from the centre of the cluster lower than R_{200}[5] virial radius of the cluster were associated with that cluster. Here, R_{200} is the radius at which the average density of a galaxy cluster is 200 times the critical density of the universe at that redshift. We found that 1 310 LSBGs from the combined catalogue and 123 LSBGs from our new sample were associated with 130 and 53 clusters, respectively. Using the redshift of the cluster provided in Xu et al. (2022), and assuming that the associated LSBG is at the same redshift as the cluster, we estimated the half-light radius of the LSBG and its projected comoving distance from the cluster centre. It should be noted that since we perform our crossmatching with only projected distances, some of the LSBGs associated with clusters could be non-cluster members that are projected along the field. However, it is unlikely to be the case for all, and since we do not have any other distance estimate for the LSBGs, we chose to adopt this method. However, it should be also noted that UDGs are not exclusively located in clusters; they can also be observed in groups (Cohen et al. 2018; Marleau et al. 2021) and

[5]We used the R_{500} values and the redshifts provided by Xu et al. (2022) to obtain the R_{200} crossmatching radius. Following Ettori & Balestra (2009), we assume $R_{200} \approx R_{500}/0.65$ where R_{500} is the radius at which the average density of a galaxy cluster is 500 times the critical density of the universe at that redshift.

even in field environments (Prole et al. 2019). In this section, we are only focusing on the LSBGs and UDGs associated with the clusters.

Among the 1,310 cluster LSBGs, we further classify 317 cluster UDG candidates based on their half-light radius ($r_{1/2} \geq 1.5$ kpc) and the central surface brightness ($\mu_0 > 24.0$mag arcsec^{-2}) in the g-band. Since we have not confirmed the physical distances to these galaxies and hence their physical sizes, they can only be regarded as UDG candidates. From here onward, when referring to UDGs in the paper, it is important to note that we are addressing UDG candidates and not confirmed UDGs. These 317 UDGs are distributed within 80 clusters making it the largest sample of clusters in which UDGs are studied. It should also be noted that Tanoglidis et al. (2021b) also identified 41 UDGs from their LSBG sample in DES by associating the 9 most overdense regions of LSBGs with known clusters. However, they did not study the properties of those 41 UDGs in detail, and the 276 UDGs among the 317 UDGs reported here are completely new. The UDGs presented here have a median $r_{1/2}$ of 2.75 kpc and μ_0 of 24.51mag arcsec^{-2}. Six of the newly identified UDGs are shown in Fig. 4.20.

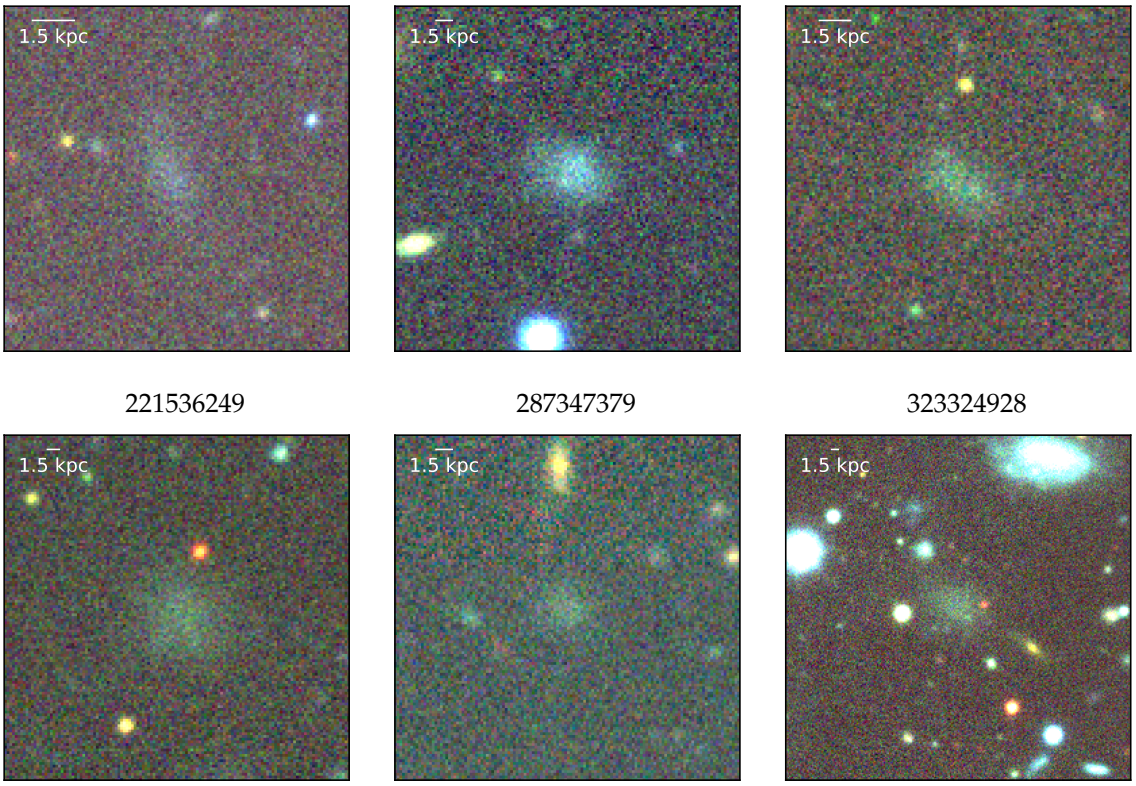


FIGURE 4.20: Cutouts of 6 confirmed new UDGs. The unique identification number (co object id) for each galaxy in DES DR1 is given below each image. The images were generated by combining the g, r and z bands using APLpy package (Robitaille & Bressert 2012), and each image corresponds to a 33.66″ × 33.66″ region of the sky with the UDG at its centre.

As seen from Fig. 4.21, the majority of the cluster UDGs (253 out of 317) are red in color ($g - i > 0.6$ mag), which is similar to the trend of cluster LSBGs (909 out of 1310). This is consistent with theoretical predictions for cluster UDGs (Benavides et al. 2023). Mancera Piña et al. (2019) have also found similar distribution for the $g - r$ color of 442 UDGs observed in 8 galaxy clusters. The joint distribution of the red and blues UDGs in

the space of $r_{1/2}$ and μ_0 is shown in Fig. 4.22. The red UDGs presented here have a median $r_{1/2}$ of 2.75 kpc and μ_0 of 24.52mag arcsec^{-2}. Similarly, the blue UDGs have a median $r_{1/2}$ of 2.78 kpc and μ_0 of 24.41mag arcsec^{-2}. Most of the red and blue UDGs have a half-light radius in the range $1.5 < r_{1/2} < 6$ kpc. However, there is a small fraction of UDGs (6 out of 317) with $r_{1/2} > 10$ kpc, which is all red and have $\mu_0 < 25.0$mag arcsec^{-2} which might be good potential candidates for the follow-up studies.

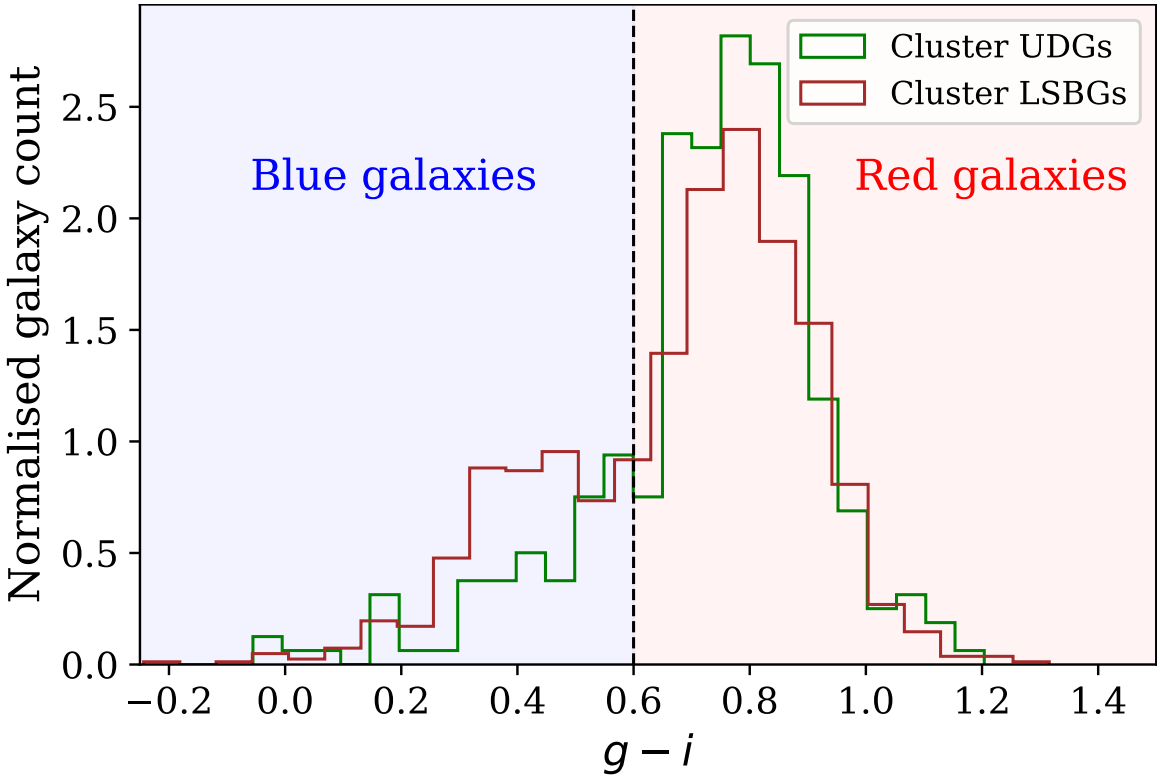


FIGURE 4.21: Color distribution of the 1,310 cluster LSBGs and 317 cluster UDGs from the DES DR1.

For all the cluster LSBGs, we can see a gradient in color as shown in Fig. 4.23, where LSBGs towards the outskirts of clusters tend to be bluer than those in the centre. This is similar to the behaviour found in Virgo cluster LSBGs from Junais et al. (2022). However, for the cluster UDGs presented in this study, the color gradient appears much weaker, almost showing a flat distribution in comparison to the LSBGs. A similar weak trend, where more blue UDGs are found towards the cluster centre, was also noted by Mancera Piña et al. (2019). On the other hand, Román & Trujillo (2017b) and Alabi et al. (2020) reported a more pronounced color trend as a function of cluster-centric distance, while La Marca et al. (2022a) did not find any significant trend. However, when directly comparing the trends in the color of UDGs in the cluster, one should keep in mind that these trends will be affected by several factors like the used bands for the color, sample size and the studied cluster, as we can see from the results in the literature, For example, our sample size (>300) is similar to the sample size of Mancera Piña et al. (2019) and have similar results whereas it is different from the findings of Román & Trujillo (2017b); Alabi et al. (2020) and La Marca et al. (2022a) which have a smaller sample size (<40).

The trend observed in the half-light radius (Fig. 4.23) for both the cluster LSBGs and UDGs is quite evident. As we move towards the outer regions of the cluster centre, both LSBGs and UDGs show an increase in size. This behaviour is in agreement with the findings of Román & Trujillo (2017b). The gradients we observe in color and size with respect

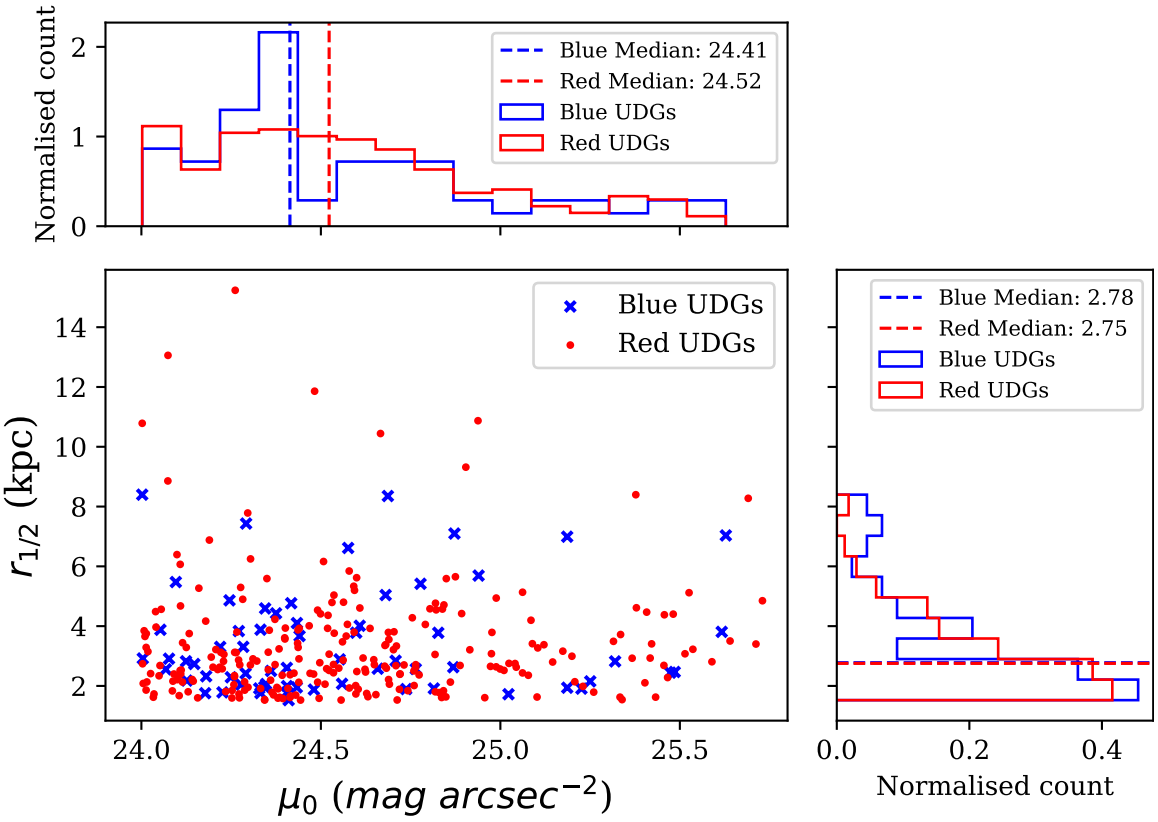


FIGURE 4.22: Joint distribution of the red (red dots) and blue (blue cross) UDGs in the space of $r_{1/2}$ and μ_0 in the g-band. The vertical lines in the histogram on the x-axis and y-axis show the median for each class.

to the cluster-centric distance are consistent with the proposed UDG formation scenarios such as the galaxy harassment (Conselice 2018), tidal interactions Mancera Piña et al. (2019), and ram-pressure stripping (Conselice et al. 2003b; Buyle et al. 2005). Such trends are also similar to what is observed for dwarf galaxies in the literature (Venhola et al. 2019a), providing further support for the argument that UDGs can be considered as a subset of dwarf galaxies (Conselice 2018; Benavides et al. 2023).

The sample of UDG candidates presented here will be the subject of the follow-up analysis. Additionally, it should be noted that all the UDGs reported here are cluster UDGs. The actual number of UDGs in the LSBG catalogue (including low-density environments) might be more than this, and thus the reported number is only a lower limit on the total number of UDGs.

4.9 Conclusions

In this paper, we explore the possibilities of using transformers in distinguishing LSBGs from artefacts in optical imaging data. We implemented four transformer models that combined the use of CNN backbone and self-attention layers to classify the labels; we call them LSBG DETR (LSBG detection transformers) models. Similarly, we have created four transformer models that directly apply attention to the patches of the images without any convolutions and these models we call LSBG vision transformers. We compared the performance of these two different architectures to the LSBG identification CNN model called DeepShadows presented in Tanoglidis et al. (2021a). We found that the transformer models performed better than the DeepShadows, and later we used the ensemble of our

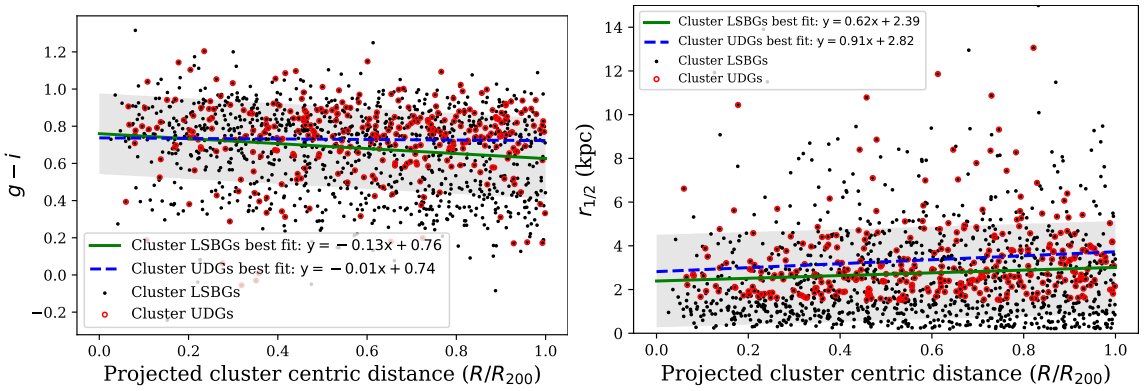


FIGURE 4.23: $g-i$ colour of the cluster LSBGs (black points) and $r_{1/2}$ as a function of the projected distance from their cluster centre (in units of the cluster radius R_{200}) is shown in the left and right panel respectively. The UDGs are marked as red hollow circles. The green line and the grey-shaded region are the linear best fit and the 1σ scatter for the cluster LSBGs, respectively. The blue dashed line is the linear best fit for the cluster UDGs.

transformer models to look for new LSBGs in the DES DR1 data that the previous searches may have missed. We follow the definition of an LSBG used by Tanoglidis et al. (2021b), i.e. we define LSBGs as galaxies having a g-band mean surface brightness $\bar{\mu}_{eff} > 24.2$ mag arcsec^{-2} and half-light radii $r_{1/2} > 2.5''$. Following this definition, we identified 4 083 new LSBGs from the DES DR1, increasing the number of identified LSBGs in DES by 17%.

Our sample selection and LSBG identification pipeline consist of the following steps:

- We preselect the objects from the DES Y3 Gold catalog based on the selection criteria described in Tanoglidis et al. (2021b) using the `SourceExtractor` parameters.
- We applied the ensemble of transformer models to this sample of preselected objects. We chose the objects identified independently by both the LSBG DETR ensemble and the LSBG ViT ensemble for a further follow-up to be inspected for being an LSBG.
- We performed a Sérsic fitting using `Galfit` and re-applied the selection cuts to further reduce the number of false positives. After this step, 4879 LSBG candidates remained to be visually inspected.
- After the visual inspection, we report the presence of 4083 new LSBGs identified by the transformer ensemble models.

Following Tanoglidis et al. (2021b), we divided the total LSBG sample into two subsamples according to their $g-i$ color. Among the 4083 new LSBGs presented here, 72% were identified as blue LSBGs, which is higher than the 67% observed in the sample presented by Tanoglidis et al. (2021b). Additionally, we also found that we have a more fraction of red LSBGs with color, $g-i > 0.8$, compared to the sample of LSBGs presented by Tanoglidis et al. (2021b). We speculate that the bias might have originated from the training set used by Tanoglidis et al. (2021b) to train the SVM model to preselect the LSBG candidate sample.

By combining the previously identified 23,790 LSBGs from Tanoglidis et al. (2021b) with the LSBGs newly identified in our work, the total number of known LSBGs in the DES is increased to 27,873. This increases the number density of LSBGs in the DES from 4.13 to 4.91 deg^{-2} for LSBGs with $\bar{\mu}_{eff}$ >24.3mag arcsec^{-2} and from 4.75 to 5.57 deg^{-2} for LSBGs with $\bar{\mu}_{eff}$ >24.2mag arcsec^{-2}. It should be stressed that this is a lower limit to the

number density, and it would increase in the future with better imaging quality and better methodology for the surveys like LSST and Euclid.

We also made an analysis of the clustering of LSBGs in DES. We found that the LSBGs tend to cluster strongly in comparison to the HSBGs from DES, which is similar to the findings by Tanoglidis et al. (2021b). Upon further examination, we observed that the strong clustering tendency observed among low surface brightness galaxies (LSBGs) primarily stems from the red LSBGs, while the behaviour of blue LSBGs resembles that of blue high surface brightness galaxies (HSBGs) with weaker clustering tendencies. Additionally, we noted a decrease in the number of red LSBGs near the centre of the galaxy cluster, resulting in a flattening of the auto-correlation function on smaller scales which is similar to the conclusions of Wittmann et al. (2017).

Additionally, we crossmatched the LSBGs with the X-ray-selected galaxy cluster catalogue from the ROSAT All-Sky Survey (RXGCC; Xu et al. 2022) to find LSBGs associated with the clusters. Using the redshift information of the clusters, we identify 317 UDGs, among which 276 are reported for the first time. We also observed a color gradient among the cluster LSBGs, where LSBGs located towards the outskirts of clusters exhibit a bluer color compared to those at the centre, similar to findings by Junais et al. (2022) in the Virgo cluster LSBGs. However, this trend is relatively weak for the cluster UDGs in our study, unlike the LSBGs. A clear trend in the half-light radius of the cluster LSBGs and UDGs as a function of the cluster-centric distance is also visible. The LSBGs and UDGs grow in size as going from the cluster centre to the outskirts. The coherent trends in the color and size are in agreement with the proposed UDG formation mechanisms such as the galaxy harassment (Conselice 2018), tidal interactions Mancera Piña et al. (2019), and ram-pressure stripping (Conselice et al. 2003b; Buyle et al. 2005), giving more support to the argument that the UDGs are a subset of dwarf galaxies (Conselice 2018; Benavides et al. 2023).

The upcoming large-scale surveys such as LSST and Euclid are expected to cover around 18 000 and 14 5000 $\deg^2$ of the sky, respectively (Ivezić et al. 2019; Euclid Collaboration et al. 2022b). Extrapolating our results on the number density of LSBGs, we are expected to find more than 100 000 and 80 000 LSBGs from LSST and Euclid, respectively. In this scenario, an improved and efficient methodology will be highly significant, and we propose that transformer models could overcome this difficulty. With the aid of transfer learning, we are planning to extend our study to HSC SSP DR3 and thus pave a pathway for the LSBG detection in LSST and Euclid.

5

DES to HSC: Detecting low surface brightness galaxies in the Abell 194 cluster using transfer learning on deep data

This chapter will be submitted to Astronomy & Astrophysics as 'DES to HSC: Detecting low surface brightness galaxies in the Abell 194 cluster using transfer learning on deep data' by Thuruthipilly H. et al. in the coming weeks. This work is done with the dedicated HSC observations of the Abell 194 cluster obtained by Jin Koda, Masafumi Yagi, Yutaka Komiyama, Yamanoi Hitomi and in collaboration with Samuel Boissier.

5.1 Introduction

Our statistical understanding of how the universe evolves is strongly biased by the objects and structures that are brighter than the surface-brightness limits of wide-area surveys. For instance, the galaxy completeness of the Sloan Digital Sky Survey (SDSS; Abazajian et al. 2009) decreases rapidly for surface brightness fainter than 24.5 mag arcsec^{-2} (e.g. Kniazev et al. 2004; Driver et al. 2005) indicating that the low surface brightness (LSB) regime is underrepresented in our catalogues. One of the key components of the LSB universe is the low surface brightness galaxies (LSBGs) characterised by their central surface brightness which is fainter than the night sky (Bothun et al. 1997). Both simulations (e.g. Martin et al. 2019) and observations (e.g. Dalcanton et al. 1997; O'Neil & Bothun 2000) indicate that the bulk of the galaxy population resides in the LSB regime. In literature, galaxies with B-band central surface brightness $\mu_0(B)$ greater than a certain threshold value are classified as LSBGs and this threshold value varies among different works, ranging from $\mu_0(B) \geq 23.0$ mag arcsec^{-2} (Bothun et al. 1997) to $\mu_0(B) \geq 22.0$ mag arcsec^{-2} (Burkholder et al. 2001).

LSBGs constitute the dominant population of the faint end of the galaxy luminosity function (Blanton et al. 2005; Martin et al. 2019) and it is estimated that the LSBGs account for a significant fraction $(30\% \sim 60\%)$ of the total number density of galaxies (McGaugh 1996; Bothun et al. 1997; O'Neil & Bothun 2000; Haberzettl et al. 2007; Martin et al. 2019). In addition, LSBGs have been found to contribute as much as 15% of the dynamical mass content of the Universe (Driver 1999; Minchin et al. 2004). Hence, LSBGs are capable of contributing significantly to our understanding of galaxy evolution (Bullock et al. 2001; de

Blok et al. 2001; Sales et al. 2020) and provide important observational constraints for cosmological models (Moore et al. 1999; Bullock & Boylan-Kolchin 2017b; Laudato & Salzano 2023).

LSBGs could be classified into several sub-classes based on their physical size, surface brightness and gas content. For instance, ultra-diffuse galaxies (UDGs) are extended LSBGs with effective radii $r_{1/2} > 1.5$ kpc and central surface brightness $\mu(g, 0) > 24$ mag arcsec^{-2}. Although the term 'UDG' was coined by van Dokkum et al. (2015a) very recently, such galaxies had been already identified in several earlier studies (Sandage & Binggeli 1984; McGaugh & Bothun 1994; Dalcanton et al. 1997; Conselice et al. 2003a). Similarly, giant LSBGs (GLSBGs) form another sub-class of LSBGs that are extremely gas-rich ($M_{HI} > 10^{10}\ M_{\odot}$), faint and extended (Sprayberry et al. 1995; Saburova et al. 2023; Junais et al. 2024). Another sub-class of LSBGs is the almost dark galaxies (ADGs) (Janowiecki et al. 2015; Leisman et al. 2017; Xu et al. 2023; Montes et al. 2024). These galaxies are typically characterized by their gas-rich nature but extremely faint optical emission making them difficult to detect in wide-field surveys such as the SDSS. The formation and evolution of these extreme sub-classes of LSBGs like UDGs, GLSBGs and ADGs are still debated and they give a robust platform to test our models of galaxy evolution and cosmology (Amorisco & Loeb 2016a; Di Cintio et al. 2017; Saburova et al. 2021; Benavides et al. 2023; Laudato & Salzano 2023; Montes et al. 2024).

A major fraction of the LSBGs had remained undetected up to the last decade and the recent advancements in digital imaging combined with deep, large sky surveys have reignited an interest in LSBGs and their detection (e.g., van Dokkum et al. 2015b; Mihos et al. 2015; Koda et al. 2015; Yagi et al. 2016; Román & Trujillo 2017a; Greco et al. 2018; Tanoglidis et al. 2021b; Bautista et al. 2023; Thuruthipilly et al. 2024b). These efforts have started to give us a glimpse of the LSBG regime and it is speculated that the LSBG population inhabits all regions of the observable universe ranging from clusters (Adami et al. 2006) to fields (Prole et al. 2019).

The new era of large-scale surveys which are both deep and wide such as the Hyper Suprime-Cam Subaru Strategic Program (HSC-SSP; Aihara et al. 2018), Euclid (Laureijs et al. 2011), and Rubin Observatory's Legacy Survey of Space and Time (LSST; Ivezić et al. 2019) is expected to uncover more LSBGs. These surveys will revolutionise our understanding of LSBGs and our knowledge of the evolution of the Universe in general.

One of the major obstacles to overcome in detecting LSBGs in the aforementioned large-scale surveys is the separation of artefacts from LSBGs. As noted by Tanoglidis et al. (2021b), these artefacts primarily consist of diffuse light from nearby bright objects, galactic cirrus, star-forming tails of spiral arms, and tidal streams. These artefacts often pass simple selection cuts to separate LSBGs from artefacts and constitute the majority of the LSBG candidate sample. Removing these contaminants is necessary, a task often achieved using semi-automated methods with a low success rate, and visual inspection, which is more precise but time-consuming.

For instance, in HSC-SSP, Greco et al. (2018) applied selection cuts on the photometric measurements from `SourceExtractor` (Bertin & Arnouts 1996) and used a galaxy modelling pipeline (based on `imfit`; Erwin (2015)) to create an LSBG candidate sample. Similarly, in dark energy survey (DES), Tanoglidis et al. (2021b) also used a pipeline that employs `SourceExtractor`, an ML algorithm and `Galfit` (Peng et al. 2002) in the respective order to create a candidate sample. However, upon visual inspection in both cases, it was found that around 50% of the shortlisted candidates were artefacts. For large-scale surveys such as HSC-SSP, LSST and Euclid these results show that the semi-automated methods would not be effective as the sheer amount of data generated will be very large and we will need pipelines with better classification accuracy to achieve meaningful results.

Recently, the advancements in deep learning (DL) have opened up many opportunities and have been widely applied in astronomy. Particularly, convolutional neural networks (CNNs) and transformers have been found to be effective in analysing astronomical data. For example, the CNNs have been used for galaxy classification (Pérez-Carrasco et al. 2019), galaxy merger identification (Pearson et al. 2022), supernova classification (Cabrera-Vives et al. 2017), classification of LSBGs (Tanoglidis et al. 2021a; Su et al. 2024) and finding strong gravitational lenses (Schaefer et al. 2018; Davies et al. 2019; Rojas et al. 2022). Similarly, transformer models have been used to detect and analyse strong gravitational lensing systems (Thuruthipilly et al. 2022; Huang et al. 2023; Jia et al. 2023; Thuruthipilly et al. 2024a), estimating cosmological parameters (Hwang et al. 2023), identification of LSBGs (Thuruthipilly et al. 2024b), representing light curves (Allam & McEwen 2021), and classifying multi-band light curves of different supernovae (SN) types (Pimentel et al. 2023).

One of the crucial requirements for creating a trained DL model is a sufficiently large training dataset that can be used to generalise the features of the data that we are trying to analyse. The recent searches for LSBGs in large-scale surveys such as SDSS (Zhong et al. 2008), HSC-SSP (Greco et al. 2018) and DES (Tanoglidis et al. 2021b; Thuruthipilly et al. 2024b) have resulted in a sufficient number of LSBGs and artefacts which could be used as training sets for creating DL models efficient in the detection of LSBGs (Tanoglidis et al. 2021a; Yi et al. 2022; Xing et al. 2023; Thuruthipilly et al. 2024b; Su et al. 2024). However, in all the previous attempts with DL models for the identification of LSBGs, the models were trained and employed for the detection of LSBGs on the same survey. In this context, a key question arises: to what extent can the knowledge acquired by a DL model trained on one survey be transferred to another survey? Specifically, can a DL model trained on the data from a survey such as DES could be effectively used to detect LSBGs in deeper data from HSC, LSST, and Euclid?

Generally, the above-mentioned questions come under the regime of transfer learning where a model trained for one task is adapted to a different task, typically by fine-tuning the model with a smaller training set (Yosinski et al. 2014). Previous efforts to transfer the knowledge of CNN models from one survey to another for tasks like galaxy morphological classification (Domínguez Sánchez et al. 2019) and classification of LSBGs (Tanoglidis et al. 2021a) have shown promise. However, the performance of the CNN drops significantly when transitioning from one survey to another necessitating fine-tuning the models to achieve satisfactory results.

To address the above-mentioned questions, here we explore the possibility of applying transfer learning to identify LSBGs in the Abell 194 cluster with the deep data we obtained from the dedicated observation with HSC. We use two different ensemble transformer models trained on the data from DES data release 1 (DES DR 1) and use the ensemble models to identify LSBGs from the HSC data which is deeper than DES DR 1 by two orders of magnitude. For consistency and comparison, we use the same transformer models used in Thuruthipilly et al. (2024b) in this work and compare the performance of the models in this work with Thuruthipilly et al. (2024b). We selected Abell 194 for our analysis since it has been covered by DES, and the small size of the cluster makes it easy to investigate the number of LSBGs that might be missed by the ensemble model. With the final sample of LSBGs and UDGs identified from this work, we also investigate the effect of the cluster environment on the morphological properties of these galaxies. Furthermore, we also investigate the presence of recent star formation in these LSBGs and UDGs with the data from the Galaxy Evolution Explorer (GALEX; Martin et al. 2005).

Throughout this paper, we adopt the following notations: the apparent magnitude is m (mag), half-light radius is $r_{1/2}$ (arcsecond), central surface brightness is μ_0, mean surface

brightness within the $r_{1/2}$ is $\bar{\mu}_{eff}$ (mag arcsec^{-2}), Sérsic index (n), axis ratio (b/a), and position angle (PA) in degrees. When specifying a particular band, we will use a subscript, such as $r_{eff,g}$ or $r_{eff,r}$ for the g and r-bands, respectively. In this paper, we adopt the cosmological parameters of $(h_0, \Omega_M, \Omega_\lambda)$ = (0.697, 0.282,0.718) following Hinshaw et al. (2013b) which corresponds to a luminosity distance of 77.7 Mpc and an angular diameter distance of 75.0 Mpc to the Abell 194 which is at redshift, $z = 0.01781$ (Girardi et al. 1998; Rines et al. 2003). In addition, an angular separation of 1 arcsecond corresponds to a distance of 0.364 kpc at the redshift of Abell 194. (Rines et al. 2003) has estimated the Abell 194 cluster to have a virial radius (R_{200}) radius of 0.9824 Mpc and a virial mass (M_{200}) of $7.6 \times 10^{13} M\odot$ (M_{200}) where R_{200} is defined as the radius at which the average density of a galaxy cluster is 200 times the critical density of the universe at that redshift.

For comparison purposes, throughout this work, we follow the LSBG definition from Tanoglidis et al. (2021b), based on the g-band mean surface brightness ($\bar{\mu}_{eff,g}$) and the half-light radii ($r_{eff,g}$). We consider LSBGs as galaxies with $\bar{\mu}_{eff,g} > 24.2$ mag arcsec^{-2} and $r_{eff} > 2.5''$. Similarly, for defining UDGs, we follow the definition on the g-band central surface brightness ($\mu_{0,g}$) and the half-light radii ($r_{eff,g}$) such that all the galaxies which have $\mu_{0,g}$ and $r_{eff,g}$ greater than 24.0 mag arcsec^{-2} and 1.5 kpc, respectively are considered as UDGs (van Dokkum et al. 2015a; Román & Trujillo 2017a).

The paper is organised as follows: Sect. 5.2 discusses the data we used to train our models and look for new LSBGs. Section 5.3 provides a brief overview of the methodology used in our study, including the models' architecture, information on how the models were trained, and the details about the visual inspection. The results of our analysis are presented in Sect. 5.4. A detailed discussion of our results and the properties of the newly identified LSBGs are analysed in Sect. 5.5. Section 5.6 concludes our analysis by highlighting the significance of LSBGs, the impact of our methodology in finding LSBGs, and the prospects with upcoming large sky surveys.

5.2 Data

5.2.1 DES

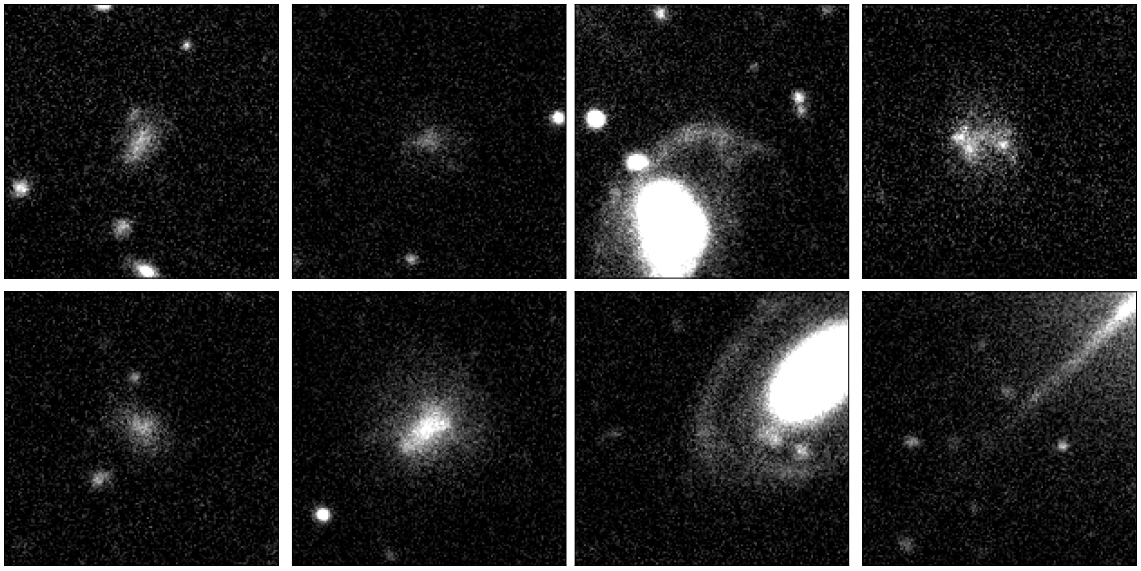
The Dark Energy Survey (DES; Abbott et al. 2018a, 2021) is a six-year observing program (2013-2019) covering $\sim$ 5000 deg^2 of the southern Galactic cap in the optical and near-infrared regime using the Dark Energy Camera (DECam) on the 4-m Blanco Telescope at the Cerro Tololo Inter-American Observatory (CTIO). The DES has observed the sky in *grizY* photometric bands with approximately 10 overlapping dithered exposures in each filter (90 sec in *griz*-bands and 45 sec in *Y*-band). The median surface brightness depth at 3σ for a $10'' \times 10''$ region for g-band is $28.26^{+0.09}_{-0.13}$ and for r-band is $27.86^{+0.10}_{-0.15}$ where the upper and lower bounds represent the 16th and 84th percentiles of the distribution over DES tiles (Tanoglidis et al. 2021b).

Training data

All of the trained, validated, and tested models in this study used the labelled dataset of LSBGs and artefacts identified from DES by Tanoglidis et al. (2021b) and extended by Thuruthipilly et al. (2024b). From the extended sample of 27 873 LSBGs (23 790 from Tanoglidis et al. (2021b) and 4 083 from Thuruthipilly et al. (2024b)), we randomly selected 18 532 LSBGs which was assigned a label 1. The catalogue for the artefacts was created based on the

publicly available dataset of artefacts [1] which consists of 20 000 artefacts. However, Thuruthipilly et al. (2024b) have shown that some of the artefacts listed in the catalogue are not artefacts and are LSBGs. After removing these artefacts from the catalogue by cross-matching them with the LSBG catalogue, we had 18 468 artefacts remaining in the list which were assigned a label 0.

We generated multi-band cutouts for each object in the flexible image transport system (FITS) format from the DES DR 1 image data in g-band and r-band. Each cutout corresponds to a 40″ × 40″ (152 × 152 pixel^2) region of the sky and is centred at the coordinates of the object (LSBG or artefact). To reduce the computational cost, we resized the cutouts from their initial size to 64 × 64 pixel^2. The cutouts of g, and r-bands were stacked together to create the dataset for training the models. Our training catalogue contains 38 500 objects, out of which 18 532 are LSBGs and 18 468 are artefacts. Before training, we randomly split the full sample into a training set, a validation set and a test set, each consisting of 32 000, 2 000, and 4 500 objects, respectively. Examples of LSBGs and artefacts used in the training set are given in Fig. 5.1.



(A) (B)

FIGURE 5.1: g- band cutouts of four examples of LSBGs (5.1a) and artefacts (5.1b) used in the training data. Each cutout of the LSBG and artefact corresponds to a 40″ × 40″ (152 × 152 pixel^2) region of the sky centred around the LSBG or artefact.

5.2.2 HSC

The Hyper Suprime-Cam (HSC) is an imaging camera covering an area of 1.77 degree^2 situated at the prime focus of the Subaru telescope (Bosch et al. 2018a). It comprises a total of 116 CCDs, featuring 104 science sensors, 4 guide sensors, and 8 focus sensors. Each CCD has dimensions of $2k \times 4k$ pixels and a pixel scale of 0.168″ per pixel (Bosch et al. 2018a; Furusawa et al. 2018; Kawanomoto et al. 2018; Miyazaki et al. 2018; Komiyama et al. 2018). The Abell 194 cluster was observed in g- and r-bands using HSC in a single pointing with dithered exposures. The integration time was approximately 48 minutes for the r-band and 73.5 minutes for the g-band. The dithering pattern was optimized to ensure that gaps between CCDs did not overlap. Each dithered integration (referred to as a "visit") began

[1] https://github.com/dtanoglidis/DeepShadows/blob/main/Datasets

with the rotator angle set to 0 degrees to maintain consistent flat patterns near the edge of the field of view. The typical seeing during the observing runs was about 1″. For easier data handling, the tract was divided into square patches of 12 arcminutes on a side (4200 × 4200 pixels), with a 17″ (100 pixels) overlap between adjacent patches, resulting in a total of 94 patches. The coverage of the Abell 194 cluster is given in Fig. 5.2.

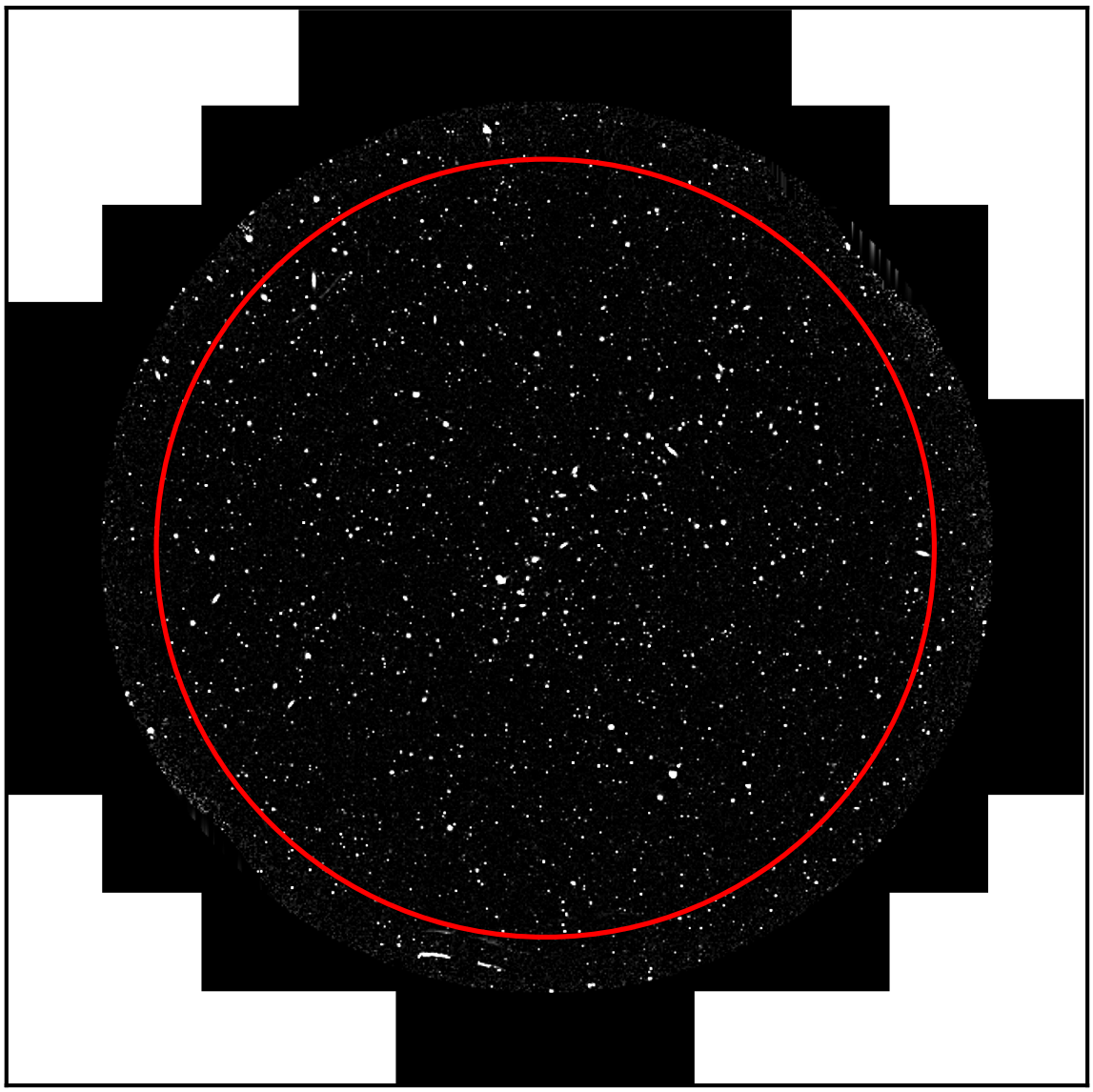


FIGURE 5.2: Coverage of the Abell 194 Cluster by HSC. The red circle marks the boundary of the virial radius of the cluster which is ∼1 Mpc (0.75 degree). The observed data covers up to 1.2 Mpc (∼ 0.8 degree) from the cluster centre.

The data were reduced using the HSC pipeline, hscPipe version 4.0.1 (Bosch et al. 2018b), which is based on software developed for the LSST project (Ivezic et al. 2008; Jurić et al. 2017), along with an additional sky-subtraction package provided by the HSC Helpdesk. Astrometric and photometric calibrations were performed using the SDSS-III DR9 catalogue (Ahn et al. 2012), which is included with the HSC pipeline and recalibrated by the HSC software team for photometric zero points against the Panoramic Survey Telescope & Rapid Response System 1 (Tonry et al. 2012; Schlafly et al. 2012; Magnier et al. 2013). The background sky was subtracted using a 512-pixel grid (approximately 87″).

We utilized fully reduced, photometrically and astrometrically calibrated, sky-subtracted images from each visit to create median-stacked images with `imcio2` (Yagi et al. 2002), minimizing bright artefacts from single exposures.

To estimate the median surface brightness depth at 3σ for a $10'' \times 10''$ region for the HSC data, we followed the procedure described in Román et al. (2020). We randomly took 1000 patches of the sky with each covering $10'' \times 10''$ region of the sky from each tile for 94 tiles (total 94 000 random patches) and fitted the counts in each path for a Gaussian profile. The σ values from each patch were used to estimate the surface brightness depth at 3σ for a $10'' \times 10''$ region using the relation (Román et al. 2020):

$$\mu_{lim}(3\sigma_{10\times10}) = -2.5 \times log\left(\frac{3\sigma}{pix \times 10}\right) + ZP \tag{5.1}$$

Here, pix is the pixel scale (0.168 arcsec/pix) of the HSC data and ZP is the zero point of the HSC data which is 27.0 mag. Using the 3σ surface brightness detection of all the patches in both g and r bands we estimated the median surface brightness depth at 3σ for a $10'' \times 10''$ region of the sky 30.48 ± 0.24 mag arcsec^{-2} in g-band and 29.70 ± 0.15 for r-band. Here the error corresponds to the median absolute deviation.

5.2.3 GALEX

GALEX is a NASA small explorer mission, imaging the sky in far-ultraviolet (FUV; 1344-1786) and near-ultraviolet (NUV; 1771–2831) bands simultaneously with a 50 cm diameter ultraviolet telescope (Morrissey et al. 2007). The telescope has a 1.25-degree field-of-view and has a full-width half maximum (FWHM) of 4.2″ and 5.3″ for the FUV and NUV bands respectively (Morrissey et al. 2007). Each pixel in the intensity map from the GALEX corresponds to 1.5″ in the angular scale.

For our analysis, we use the intensity maps of all the available GALEX data within a 2-degree search radius of the Abell 194 cluster from the MAST database[2]. The intensity maps were coadded using their corresponding exposure maps with `reproject` python package. The final co-added image has a variable exposure along the field of the Abell 194 cluster with a median exposure time of 4274 ± 1232 seconds in the NUV and 3943 ± 1393 seconds in the FUV band where the error denotes the median absolute deviation.

5.3 Methodology

5.3.1 Transformer models

Here we use two types of transformer models which we call LSBG detection transformers (LSBG DETR) and LSBG Vision transformers (LSBG ViT). Both architectures have been used for the identification of LSBGs from DES DR 1 by Thuruthipilly et al. (2024b) and we use the same models for the identification of LSBGs in Abell 194 with HSC data for comparing the results from this work to the results from Thuruthipilly et al. (2024b).

LSBG detection transformer (LSBG DETR)

We have created four LSBG DETR models similar to the architecture described in Sec. 2.5.2. The hyperparameters of the LSB DETR models are presented in Table 5.1.

[2] https://archive.stsci.edu/

LSBG vision transformer (LSBG ViT)

We have created four LSBG ViT models similar to the architecture described in Sec. 2.5.3. The hyperparameters of the LSB ViT models are presented in Table 5.1.

TABLE 5.1: Table showing the name of the model, size of the image patches (s), number of heads (h), number of transformer encoder layers (T) and the number of epochs taken to train (e) the models in chronological order of creation.

Model Name	s	h	T	e
LSBG DETR 1	-	8	4	110
LSBG DETR 2	-	12	4	87
LSBG DETR 3	-	12	4	128
LSBG DETR 4	-	8	4	168
LSBG ViT 1	4	12	4	82
LSBG ViT 2	4	12	8	57
LSBG ViT 3	6	12	4	87
LSBG ViT 4	6	16	8	58

5.3.2 Training

All of the LSBG DETR and LSBG ViT models were trained with an initial learning rate of $\alpha = 10^{-4}$. We used the exponential linear unit (ELU) function as the activation function for all the layers in these models (Clevert et al. 2016). We initialise the weights of our model with the Xavier uniform initialiser (Glorot & Bengio 2010b), and all layers are trained from scratch by the ADAM optimiser with the default exponential decay rates (Kingma & Ba 2015). We have used the early stopping callback from `Keras` [3] to monitor the validation loss of the model and stop training once the loss was converged. The models LSBG DETR 1, 2, 3 and 4 were trained for 110, 87, 128 and 168 epochs, respectively. Similarly, the LSBG ViT 1, 2, 3, and 4 were trained for 82, 57, 87 and 58 epochs, respectively. In the spirit of reproducible research, our code for LSBG DETR and LSBG ViT is publicly available [4].

5.3.3 Transfer learning

Transfer learning has been found to be very useful and successful in training large natural language models and computer vision models. Given the need for better machine learning models and techniques in the era of big data, the astronomy community has also delved into the potential of transfer learning. For instance, Ackermann et al. (2018) used a CNN model that has been trained on pictures of everyday objects (i.e. ImageNet data set, Deng et al. (2009)) to be retrained for the detection of galaxy mergers. Similarly, Wei et al. (2020) and Hannon et al. (2023) have used DL models pre-trained on the ImageNet data set for the classification of star clusters.

One fact that should be considered when applying transfer learning from one survey to another is the difference in the instruments used in these surveys. For example, two surveys with different zero points and pixel scales will have different pixel values for the same astronomical object in counts/second units. Hence, a model trained on one survey could not be directly used on another survey and needs some standardisation. One of the common trends is to re-scale the input image within a specific range (i.e. 0 to 1 or -1 to 1).

[3] https://keras.io/api/callbacks
[4] https://github.com/hareesht23/

This also ensures that the input values are not very small which would compromise the performance of the DL model (de Amorim et al. 2023).

An alternative way is to convert the pixel values of the images from counts/seconds to their surface brightness units. Hence, the average pixel values over a region remain the same for different surveys and could standardise the image data of the same wavelength for two different surveys. In this work, before training and testing our transformer models, we convert each pixel value to its surface brightness (μJy arcsec^{-2}) for DES DR 1 and HSC data, respectively. Here, during the conversion, the brightness of each pixel is divided by its pixel area to remove the dependence on the different pixel scales of each instrument. However, it is important to note that this standardization of data does not address the difference in point spread function (PSF) values between different surveys. It should be noted that only for the application of the ML models do we apply this conversion, and all the measurements and analysis are done on the original data.

5.3.4 Metrics for comparing models

The metrics used to compare the models are accuracy, AUROC, TPR and FPR which are described in Sect. 2.7. I refer the reader to Sect. 2.7 for a detailed discussion of these metrics.

5.3.5 Object detection and preselection of LSBG candidates in Abell 194

Due to the inherent low surface brightness, the LSBGs are generally underrepresented in galaxy catalogues. Conventionally, to create a complete catalogue of all the LSBGs from a dataset, the initial step is to create a catalogue of all the objects in the data and subsequently apply selection criteria to refine the candidate sample. To detect the objects, we use `SourceExtractor` (Bertin & Arnouts 1996) on the image patches. As mentioned in Sect. 5.1, here we consider all the galaxies that have a $\bar{\mu}_{eff} > 24.2$ mag arcsec^{-2} and $r_{1/2} > 2.5''$ as LSBGs. Hence, we apply some preselections with parameters estimated from `Sextractor` which are described below to create a preliminary sample of LSBGs.

As the first step, we use the r-band images for the detection and estimate the morphological parameters of the detected objects in the g band for the subsequent selection of LSBG candidates. We set the `DETECT_THRESH` as 1.5σ and `DETECT_MINAREA` as 49 pixels to detect all the objects that have a signal-to-noise ratio greater than or equal to 1.5 and span over at least 49 pixels ($\sim$ 1.4 arcsec^2). For comparison, it should be noted that we are looking for LSBGs which have an $r_{1/2} > 2.5''$, corresponding to an area greater than 6 arcsec^2.

We further apply selection cuts on the output catalogue based on the $r_{1/2}$, $\bar{\mu}_{eff}$ and the q as measured by the `Sextractor` to reduce the LSBG candidate sample. Since we are looking for LSBGs that are extended, we removed the bright and point objects from the full catalogue. For removing the bright objects we apply selection criteria that the LSBG candidate should have a $\bar{\mu}_{eff}$ in the g-band within the range 24.0 to 31 mag arcsec^{-2}. The upper limit to $\bar{\mu}_{eff}$ is 0.2 magnitudes brighter than the LSBG definition we adopted in this work to ensure that our conditions are conservative and we do not miss any faint sources that satisfy our LSBG definition. The lower limit to $\bar{\mu}_{eff}$ is equivalent to the 3σ the surface brightness detection limit of the HSC data. Additionally, the measured $r_{1/2}$ in the g-band was restricted to be in the range $2'' < r_{1/2} < 20''$. A relaxed lower limit for $r_{1/2}$ was chosen to maximise the number of extended LSBGs that will remain after the size selection. Finally, we restricted the axis ratio b/a (`B_IMAGE/A_IMAGE`) in the range of $0.3 < b/a \leq 1.0$ in order to remove artefacts such as the highly elliptical diffraction spikes.

It should be noted that the selection applied above is necessary to reduce the number of FPs. For instance, consider a classifier model with a TPR of 100% and an FPR of 0.1%. When

testing this model on an imbalanced dataset, such as all objects from DES DR1 containing approximately 3×10^8 objects, of which only around 3×10^5 are LSBGs, there is a statistical expectation that the model will correctly identify all 3×10^5 LSBGs (since TPR = 100%). However, based on the FPR (0.1%), it is also anticipated that the model will misclassify an artefact as LSBG for every 10,000 artefact. Hence, if the test dataset is extremely biased such as all the LSBGs and non-LSBGs in DES, it would result in too many false positives. (Petrillo et al. 2019; Petrillo et al. 2019; Rojas et al. 2022; Li et al. 2020; Thuruthipilly et al. 2024a; Grespan et al. 2024). Hence selection cuts are necessary for the reliable performance of any ML model on large datasets before applying.

Furthermore, we also crossmatch the shortlisted LSBG candidates with SDSS data release 16 (SDSS DR16 Ahumada et al. 2020) to find if any of the detected sources are foreground or background sources with respect to the Abell 194 cluster. We remove these samples from our analyses so not to be contaminated by these sources.

For each object that remained in the catalogue, we generated multi-band cutouts in g and r-bands in the flexible image transport system (FITS) format. Each cutout corresponds to a 40″ × 40″ (238 × 238 pixels) region of the sky and is centred at the coordinates of the object as similar to the training data from DES. Since some of the objects might be located near the edge of the patches, for these objects we co-added the regions from nearby patches which were overlapping with a 40″ × 40″ (476 × 476 pixels) centred around the object to ensure that no light from the object is missed in the cutout. However, it should be noted that since the pixel scale of the DES was different 80″ × 80″ region in DES corresponded to 152 × 152 pixels. We resized the cutouts from their initial size to 64 × 64 pixels to match the input size of the ensemble models. The cutouts of g, and r-bands were stacked together respectively to be tested with the ensemble models.

5.3.6 Masks

When compact objects are on top of an LSBG, it will affect the Sérsic profile fitting (discussed in Sect. 5.3.7) and the creation of accurate masks that only preserve the light from the LSBG is necessary. Using the fact that the LSBGs we have defined here are faint and extended, we mask all the objects that are brighter or smaller in the cutouts. For the masking procedure, we followed the procedure described Bautista et al. (2023). We run the `SourceExtractor` on each cutout three times by adjusting the control parameters to detect different types of contaminants such as the bright sources, faint sources lying outside the LSBG, and faint sources lying on top of the LSBG. The `CHECKIMAGE` image that contains only the detected objects is used to create the masks for each kind and all three masks are combined to create the final mask.

The first step is to remove the bright objects from the cutouts. In `SourceExtractor`, the bright objects can be removed by setting a higher detection threshold without any limits on the size. For removing the bright objects we set `DETECT_THRESH` as 23.0 mag arcsec^{-2} and masked all the bright pixels. The second step is to remove the faint small objects that are far from the central LSBG. The smaller objects in isolation can be detected by setting a maximum area threshold (`DETECT_MAXAREA`) of 200 pixels at a detection threshold of 27.5 mag arcsec^{-2}.

To find the smaller objects on top of LSBGs, we use the unsharp masking technique. For this, we smooth the original g-band image by convolving it with a Gaussian kernel and subtracting the smoothed image from the original. After the subtraction, the features with sizes on the order of, or smaller than, the smoothing kernel will remain in the subtracted images. Running the `SourceExtractor` on the subtracted image could detect all the small objects that are even on top of the LSBGs and could be removed. Hence, by choosing an appropriate value for the FWHM, the small objects on top of the LSBGs could be removed.

However, it should be noted that, in this step depending on the size of the LSBGs and also the size of the foreground objects, the FWHM of the Gaussian kernel and the DETECT_MAXAREA (maximum area threshold for a source to be detected) needed to detect the foreground objects would be different. The majority of the masks were created with FWHM = 2.0″ and DETECT_MAXAREA = 200. However, setting a universal value for these parameters could result in bad masks for some LSBGs. By trial and error, we found out the combinations of FWHM = [2.0″, 1.25″, 2.50″] and DETECT_MAXAREA = [200, 100] were used to create the accurate masks for $\sim$ 40% of LSBGs. The created mask could successfully mask the contaminants that are on top of the LSBG and preserve the light from the LSBG.

5.3.7 Sérsic fitting

All the astronomical objects from Abell 194 data classified as LSBGs by either LSBG DETR or LSBT ViT ensemble models were subjected to a single component Sérsic fitting with Galfit (Peng et al. 2002). This was done to re-evaluate the $\bar{\mu}_{eff}$ and $r_{1/2}$ values of our LSBG candidates. We opted for a single-component Sérsic fitting approach to align with the LSBG search methodology of Tanoglidis et al. (2021b) and Thuruthipilly et al. (2024b). However, it should be noted that Sérsic fitting may not always capture the complete light profile of a galaxy.

We used the magnitude (MAG_AUTO) and radius (FLUX_RADIUS) values we obtained from running the SourceExtractor as an initial guess for running Galfit. We also used A_IMAGE and B_IMAGE estimated from the SourceExtractor to calculate the initial guess for the axis ratio b/a (B_IMAGE/A_IMAGE) and was allowed to vary in the range of $0.3 < q \leq 1.0$. The Sérsic index (n) was initialised at 1 and was allowed to vary only within the range of $0.2 < n < 5.0$. The position angle was initialised at 90 deg and was allowed to vary without any restrictions to find the optimum angle. As each cutout is centre around the LSBG candidate, we initialised the X and Y positions as the centre of the cutout which is at X = 119 pixels and Y = 119 pixels. Furthermore, we allowed the centre of the LSBG candidate to be varied $\pm$5 pixels for both X and Y.

Prior to the Sérsic fitting, we subtract the local sky from each cut-out since an error in the sky determination will greatly affect the Sérsic parameters estimated for faint sources (Bautista et al. 2023). The local sky is estimated as 2.5$\times$ median - 1.5$\times$ mean of the flux from the cutout of each LSBG candidate if the median is less than the mean. If the median is greater than the mean, then the mean is treated as the local sky background (Da Costa 1992). When calculating the total flux from each cutout, a circular region of 10″ with an LSBG candidate at the centre is excluded to avoid the light from the LSBG. An example of an LSBG fitted with Galfit is shown in Fig. 5.3.

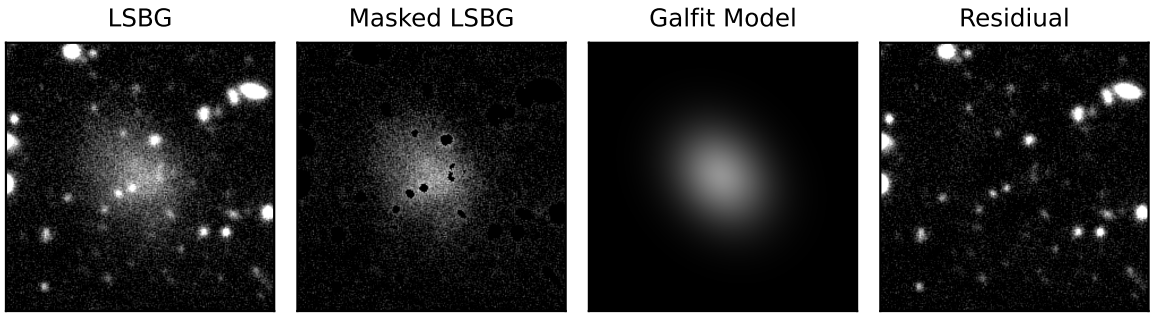


FIGURE 5.3: Image of an LSBG (RA = 21.82879, DEC=-1.81381) as observed in HSC in the g-band, the image with all the objects other than LSBG masked, corresponding Sérsic model fitted by Galfit and the residual are shown respectively from the left to the right.

A similar fitting procedure was done for both the g-band and r-band images of our sample. After the fitting, we excluded all the sources with poor/failed fits with a reduced $\chi^2 > 3$. We have also excluded the cases where the estimated n, q, `X` and `Y` values do not converge and are on the edge of the range specified above. For the remaining galaxies, we re-applied our g-band sample selection criteria of $\bar{\mu}_{eff} > 24.2$ mag arcsec^{-2} and $r_{eff} > 2.5''$. The $\bar{\mu}_{eff}$ and μ_0 values were calculated using the relations (Graham & Driver 2005):

$$\bar{\mu}_{eff} = m + 2.5log_{10}(2\pi q r_{1/2}^2), \tag{5.2}$$

and

$$\mu_0 = \bar{\mu}_{eff} + 2.5log_{10}\left(\frac{n}{b^{2n}}\Gamma(2n)\right). \tag{5.3}$$

where $\bar{\mu}_{eff}$ is the mean surface brightness within the effective radius, m is the total magnitude, q is the axis ratio, r_{eff} is the half-light radius in a specific band estimated from `Galfit`, μ_0 is the central surface brightness, n is the Sérsic index and Γ is the complete gamma function. For all our measurements, we also applied a foreground galactic extinction correction using the Schlegel et al. (1998) maps normalised by Schlafly & Finkbeiner (2011) and a Fitzpatrick (1999) dust extinction law. All the reported values from here onwards have been corrected for corrected for the Galactic extinction.

5.3.8 Visual inspection

Only the candidates identified independently by LSBG DETR and LSBT ViT ensemble models and passed the selection criteria for being an LSBG with the updated parameters from the `Galfit` were considered for visual inspection. This refined sample was subjected to visual inspection by two authors independently. Candidates identified as LSBG by both authors were treated as confident LSBGs, and candidates identified as LSBG by only one author were reinspected together to make a decision. Since visual inspection is time-consuming, we only resorted to it at the last step to reduce the number of candidates shortlisted for visual inspection.

To aid in visual inspection, we used the colour images of the LSBG candidate downloaded from the DESI Legacy Imaging Surveys Sky Viewer (Dey et al. 2019) as well as the g-band images from the HSC. Furthermore, the g-band Sérsic models from `Galfit` were also inspected visually to ensure the quality of the fits. Each candidate was then categorised into three classes based on the `Galfit` model fit and the image: LSBG, non-LSBG (Artifacts), or misfitted LSBGs. If the model of the galaxy was fitted correctly and the candidate showed LSBG features, it was classified as an LSBG. If the candidate shows LSBG features but does not fit correctly, we classify it as a misfitted LSBG and was fitted again with different initial parameters or with better masks. Finally, if the candidate does not have features of an LSBG, we classify it as an artefact or non-LSBG.

5.3.9 Aperture photometry

As mentioned above, Sérsic fitting cannot always capture the complete light profile of a galaxy and the magnitude estimated from a Sérsic profile may not exactly represent the correct apparent magnitude of the galaxy. Hence we also perform aperture photometry using `photutils`[5] for all the LSBGs to estimate the apparent magnitude inside an aperture of size 8″($\sim$ 2.9 kpc) centre around the LSBGs. The aperture photometry is performed for g and r-bands and also for NUV and FUV bands from the GALEX.

[5] https://photutils.readthedocs.io/en/stable/

The masks used during Sérsic fit are also used for the aperture photometry. However, since the NUV and FUV data has a resolution of 5″ which is larger than the size of most LSBGs, we unmasked the point objects that were masked on top of the LSBGs to not unnecessarily mask the data. These masks were created using the $g-$band images and were resized to match the pixel scale of the GALEX data. For NUV and FUV photometry the error from the sky background is estimated by choosing an annulus with an inner radius of 10″ and an outer radius of 20″ and estimating the sky background from this region. Only sources with $S/N > 3\sigma$ were considered as a confident detection and for the cases with $S/N < 3\sigma$ the 3σ value was used to report the upper limit for the NUV and FUV magnitude. The aperture magnitudes were corrected for foreground Galactic extinction similarly as discussed in Sect. 5.3.7.

5.4 Results

5.4.1 Model performance on the testing set

As mentioned in the section 5.3.1, we have developed four of each transformer models, LSBG DETR and LSBG ViT, each with different hyperparameters. Each model is a regression model, predicting the probability of an input being an LSBG. To improve the combined performance of the transformer models, we create an ensemble for LSBG DETR and LSBG ViT. Such that the outputs from the 4 models are averaged to predict the probability of an input being an LSBG. We set the classification threshold at 0.5, meaning inputs with a predicted probability greater than or equal to 0.5 are classified as LSBGs. The performance of all the modes is listed in Table 5.2 where the architecture, accuracy and AUROC of all the models on the test dataset from DES.

TABLE 5.2: Table comprising the architecture, accuracy, true positive rate (TPR), false positive rate (FPR) and AUROC of all the models in chronological order of creation.

Model name	Accuracy (%)	TPR	FPR	AUROC
LSBG ViT 1	95.58	0.96	0.05	0.9908
LSBG ViT 2	95.48	0.96	0.05	0.9906
LSBG ViT 3	95.58	0.97	0.06	0.9906
LSBG ViT 4	95.14	0.96	0.05	0.9895
LSBG ViT Ensemble	**95.62**	**0.96**	**0.05**	**0.9911**
LSBG DETR 1	95.68	0.96	0.04	0.9893
LSBG DETR 2	95.36	0.95	0.04	0.9887
LSBG DETR 3	95.48	0.96	0.05	0.9891
LSBG DETR 4	95.54	0.97	0.06	0.9904
LSBG DETR Ensemble	**95.62**	**0.96**	**0.05**	**0.9903**

On the test dataset created from the DES, both LSBG DETR and LSBG VIT models were able to achieve almost the same accuracy (95.62%), TPR (0.96), FPR (0.05) and AUROC (0.99) scores. However, it should be noted that the same metrics do not mean that the models are identical and give the same prediction probability. For instance, Fig. 5.4 shows the probability predictions of LSBG DETR and LSBG ViT ensemble models. Here, the LSBG DETR models assign a high probability value to an LSBG and a lower probability value to an artefact compared to LSBG ViT models.

The TPR and FPR reported in Table 5.2 are computed using a threshold of 0.5 to distinguish between LSBGs and artefacts. The variation of TPR and FPR as a function of the

threshold is shown in Fig. 5.5, representing the receiver operating characteristic (ROC) curve of the model. Fig. 5.5 indicates that both ensemble models can detect more than 90% of LSBGs even when with a relatively high threshold, such as 0.75.

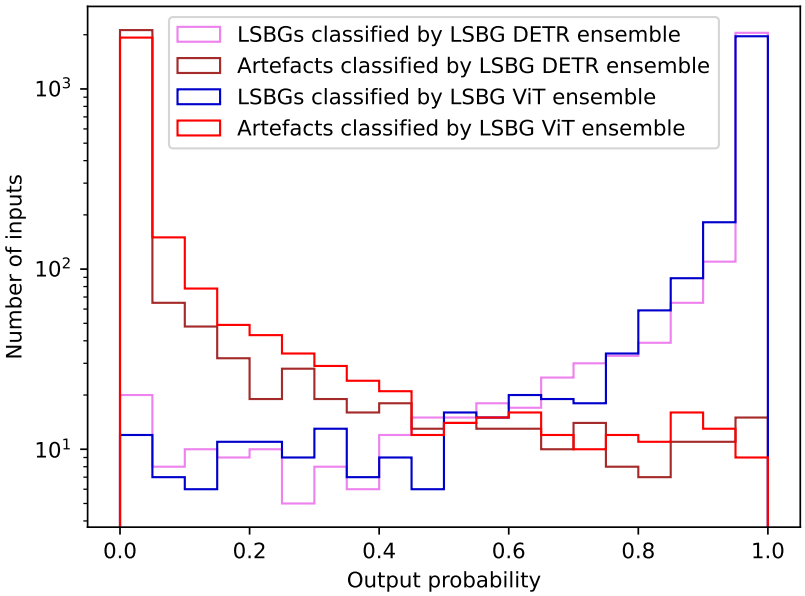


FIGURE 5.4: The probability distribution of LSBG DETR and LSBG ViT ensemble classifier models

5.4.2 Search for LSBGs with HSC in Abell 194 cluster

As mentioned in Sect. 5.3.5, we use `SourceExtractor` to identify all the objects in the dataset. By running the `SourceExtractor` patch by patch with the parameters specified in Sect. 5.3.5, we identified 170 328 sources. It should be noted that not all the detected sources are astronomical objects, and a large fraction of these sources might be instrumental or physical artefacts. To reduce the sample size of the object catalogue, we apply the selection criteria as mentioned in Sect. 5.3.5. After applying the selection cuts to $r_{1/2}$, $\bar{\mu}_{eff}$, and q, we have 991 sources detected from the Abell 194, which could be a potential LSBG. It should be noted that some of the LSBGs might be missed because of the preselection, and a more relaxed preselection compared to the definition of the LSBG is preferred to retain as many LSBGs as possible.

After applying the preselection as described in section 5.3.5, we have a crude candidate catalogue of LSBGs. Further, we cross-matched the crude catalogue with SDSS DR 16 catalogue Ahumada et al. (2020) to remove the foreground and background galaxies compared to the Abell 194 cluster. We found 14 galaxies with known redshift and they were removed, successively reducing the size of the crude catalogue to 977 objects. The remaining objects in this catalogue could be an LSBG, an artefact, or a non-LSBG (a faint galaxy but not faint enough to be classified as an LSBG as per the definition that we use in this work).

To separate faint galaxies from artefacts, we apply our LSBG DETR and LSBG ViT models separately to all the objects that passed the selection criteria. LSBG DETR models identified 258 LSBG candidates, and LSBG ViT models identified 261 LSBG candidates

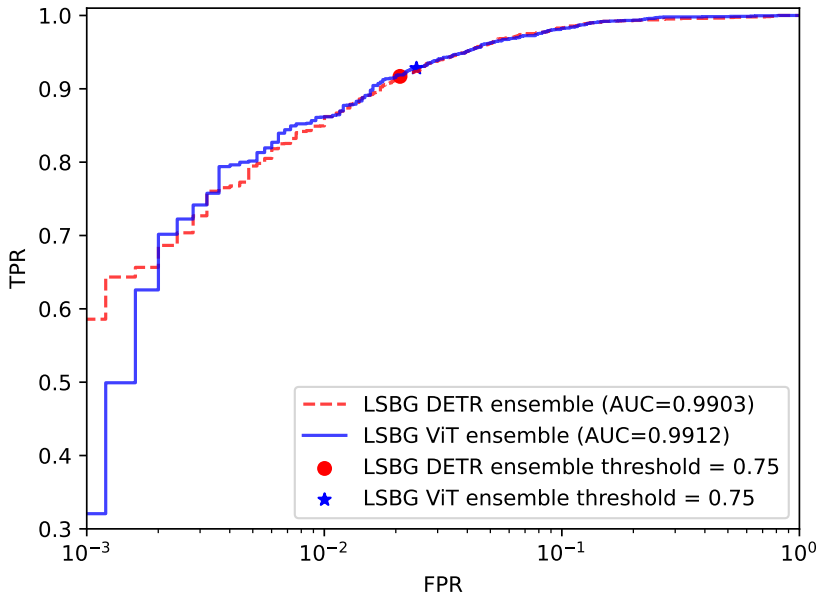


FIGURE 5.5: The receiver operating characteristic (ROC) curve of the LSBG DETR and LSBG ViT models.The red and blue lines represent the variation of FPR and TPR as a function of the threshold for LSBG DETR and LSBG ViT ensembles, respectively. The red and blue points mark the TPR and FPR for a threshold = 0.75.

independently and only 234 objects were classified as LSBG candidates by both models. To maximise the number of LSBGs, an input classified as an LSBG by either one of the models is treated as an LSBG candidate and passed for the subsequent analysis.

There were 272 sources classified by either one of the ensemble models as LSBG candidates and all of them underwent single-component Sérsic profile fitting using `Galfit`. Following the Sérsic profile fitting, we reapplied the selection criteria to screen the LSBG candidates. We applied selection cuts to the n, q, and centre positions X and Y of the fitter parameters as described in Sect. 5.3.7. These criteria were used to eliminate any poor fits and artefacts, if present. Two of the authors independently visually inspected the masks and fitted galaxy profiles to ensure the quality of the fit. After the `Galfit` and visual inspection, there were 159 LSBGs in the data.

To estimate the number of the LSBGs missed by our model, we also repeated the `Galfit` and visual inspection of all the 705 objects rejected by the ensemble models. This was done to have an estimate of the number of false negatives (FN) predicted by the model and to minimize FN occurrences during transfer learning in the future. We found that our model missed 12 LSBGs making the total number of LSBGs in around the Abell 194 as 171. Thus our model was able to achieve $\sim$ 93% TPR on the HSC dataset which is comparable to the TPR achieved by the ensemble models on the test data set from DES. The schematic diagram showing the sequential selection steps used to create LSBGs and UDGs in Abell 194 is shown in Fig. 5.6. A sample catalogue comprising the properties of the identified LSBGs in this work is shown in Table 5.3. Six examples of LSBGs and UDGs identified from our study are shown in Fig. 5.7.

Girardi et al. (1998) and Rines et al. (2003) have estimated the comoving radius of Abell 194 as 0.69 h^{-1} Mpc. Plugging in the cosmological parameters, we get $h(0.01781) = 0.7023$ and we find that the virial radius, r_{200} of Abell 194 is 0.9824 Mpc. We find that 12 of the 171

TABLE 5.3: Sample of LSBGs identified in this work. 'NAME' is the unique object ID associated with each LSBG whereas the 'RA' and 'DEC' give the sky coordinate of the LSBG. Columns 'n', 'b/a', 'PA','$log(\Sigma_{star})$' and '$log(M_{star})$' represent the Sérsic index, axis ratio, stellar mass density and stellar mass of the LSBGs, respectively. The columns '$r_{eff,g}$', 'g', '$\bar{\mu}_{\mathrm{eff},g}$', '$\mu_{0,g}$' and $\chi^2_{\nu g}$ represents the half-light radius, magnitude, mean surface brightness within the half-light radius, central surface brightness and reduced chi-square value of g-band estimated using `Galfit`. Similarly, the columns '$r_{eff,r}$', 'r', '$\bar{\mu}_{\mathrm{eff},r}$', '$\mu_{0,r}$' and $\chi^2_{\nu r}$ represents the half-light radius, magnitude, mean surface brightness within the half-light radius, central surface brightness and reduced chi-square value of r-band estimated using `Galfit`. The final column 'DES' labels the LSBG as 1 if they are reported in Tanoglidis et al. (2021b) or Thuruthipilly et al. (2024b) and 0 if it is not reported.

NAME	RA (deg)	DEC (deg)	n	b/a	PA (deg)	$log(\Sigma_{star})$ $(M_\odot kpc^{-2})$	$log(M_{star})$ $(M_\odot)$	$r_{eff,g}$ ($''$)	g (mag)	$\bar{\mu}_{\mathrm{eff},g}$ $(\frac{\mathrm{mag}}{\mathrm{arcsec}^2})$	$\mu_{0,g}$ $(\frac{\mathrm{mag}}{\mathrm{arcsec}^2})$	$\chi^2_{\nu g}$	$r_{eff,r}$ ($''$)	r (mag)	$\bar{\mu}_{\mathrm{eff},r}$ $(\frac{\mathrm{mag}}{\mathrm{arcsec}^2})$	$\mu_{0,r}$ $(\frac{\mathrm{mag}}{\mathrm{arcsec}^2})$	$\chi^2_{\nu r}$	DES
J01:22:31.599-01:37:29.491	20.6317	-1.6249	0.98	0.44	39.5	6.7	7.4	3.7	21.4	25.3	24.2	1	3.8	20.7	24.8	23.7	1.1	0
J01:22:44.406-01:21:05.462	20.685	-1.3515	0.73	0.47	-68.7	6.7	7.3	3.4	20.7	24.5	23.8	1.1	3.2	20.4	24.1	23.4	1.1	0
J01:22:57.659-01:03:57.566	20.7402	-1.066	1.09	0.74	-35.6	6.8	7.4	2.7	21.1	24.9	23.6	1	2.5	20.6	24.2	22.9	1	0
J01:23:09.271-01:28:26.721	20.7886	-1.4741	0.98	0.6	67.3	6.4	7.3	3.8	21.6	25.9	24.8	1	3.8	21	25.3	24.2	1	1
J01:23:22.943-01:17:38.620	20.8456	-1.2941	0.69	0.65	-59.3	6.4	7	3	22.3	26.3	25.7	1	2.9	21.7	25.5	24.9	1	0
J01:23:27.773-00:55:21.039	20.8657	-0.9225	0.62	0.38	-11.9	6.8	7.1	2.8	21.2	24.4	23.9	1.1	2.5	20.9	23.9	23.4	1	0
J01:23:31.637-01:23:25.319	20.8818	-1.3904	1.93	0.94	-1.5	7.7	8.1	3	20.6	24.9	22.1	1.1	1.9	19.5	22.8	20	1.2	0
J01:23:44.851-01:37:27.984	20.9369	-1.6244	0.89	0.89	-65.9	6.5	7.5	3.6	21.1	25.8	24.8	1	3.6	20.5	25.2	24.2	1.1	1
J01:23:47.268-01:17:26.726	20.947	-1.2908	0.76	0.76	77.1	6.4	7.3	3.9	21.3	25.9	25.2	1	3.9	20.7	25.4	24.7	1.1	1
J01:23:59.643-01:24:43.713	20.9985	-1.4121	0.88	0.73	50.3	6.8	7.7	4	20.3	24.9	24	1	3.8	19.7	24.3	23.4	1.1	1
J01:24:00.601-01:35:15.808	21.0025	-1.5877	1.14	0.87	70.5	7	7.8	3.4	20.1	24.6	23.2	1.1	3.2	19.6	23.9	22.6	1	1
J01:24:02.742-01:49:30.931	21.0114	-1.8253	0.87	0.97	74.5	6.8	7.7	3.2	20.6	25.1	24.2	1	3.2	20	24.5	23.6	1	1
J01:24:06.822-01:42:08.144	21.0284	-1.7023	0.7	0.63	64.1	6.4	7.4	4.5	21.3	26.1	25.5	1	4.5	20.7	25.5	24.8	1	0
J01:24:09.443-01:34:51.826	21.0393	-1.5811	1.16	0.72	-30.8	7	7.7	3.2	20.2	24.4	23	1.1	2.9	19.7	23.7	22.3	1.1	0
J01:24:09.514-01:37:19.400	21.0396	-1.6221	0.92	0.55	-43.3	6.7	7.4	3.5	21.2	25.3	24.3	1	3.4	20.7	24.6	23.6	1	1
J01:24:10.378-01:24:03.853	21.0432	-1.4011	0.92	0.5	59.1	6.9	7.4	3.2	20.5	24.2	23.2	1.2	3	20.2	23.8	22.8	1.1	0
J01:24:12.524-01:15:53.605	21.0522	-1.2649	0.78	0.78	-51.8	6.9	7.6	3.1	20.4	24.5	23.8	1.1	3	19.9	24	23.3	1	1
J01:24:16.553-01:24:54.715	21.069	-1.4152	0.99	0.89	-38.6	6.3	7	2.9	21.8	26	24.8	1	2.6	21.4	25.3	24.2	1	0
J01:24:17.177-01:30:07.435	21.0716	-1.5021	0.84	0.51	46	6.9	7.5	3.2	21.1	24.9	24	1	3.3	20.5	24.3	23.4	1.1	0
J01:24:19.742-01:36:32.984	21.0823	-1.6092	0.59	0.85	57.1	6.7	7.4	2.9	20.6	24.7	24.3	1.1	2.9	20.3	24.4	23.9	1.1	0
J01:24:24.290-01:41:09.630	21.1012	-1.686	0.71	0.8	-20.5	6.2	7.1	3.6	21.7	26.3	25.6	1	3.4	21.3	25.7	25	1	0
J01:24:24.399-01:13:30.857	21.1017	-1.2252	0.94	0.57	-26.9	6.7	7.8	5.4	20.2	25.2	24.2	1	5.3	19.6	24.7	23.7	1	1
J01:24:24.479-01:25:08.939	21.102	-1.4191	0.73	0.94	-89.8	6.4	7.1	3	21.6	25.9	25.2	1	2.8	21.2	25.3	24.6	1	0
J01:24:26.226-01:24:45.783	21.1093	-1.4127	1.13	0.42	-24.5	7	8	6	19.4	24.3	23	1.4	5.5	19	23.7	22.4	1.1	0
J01:24:28.295-01:25:03.997	21.1179	-1.4178	0.75	0.6	-49.6	6.6	7.5	4	21	25.4	24.7	1	3.8	20.4	24.8	24.1	1	1
J01:24:35.281-01:31:10.590	21.147	-1.5196	1	0.6	-88.6	6.5	7.2	3.4	21.6	25.7	24.6	1.1	3.1	21.1	25	23.9	1.1	0

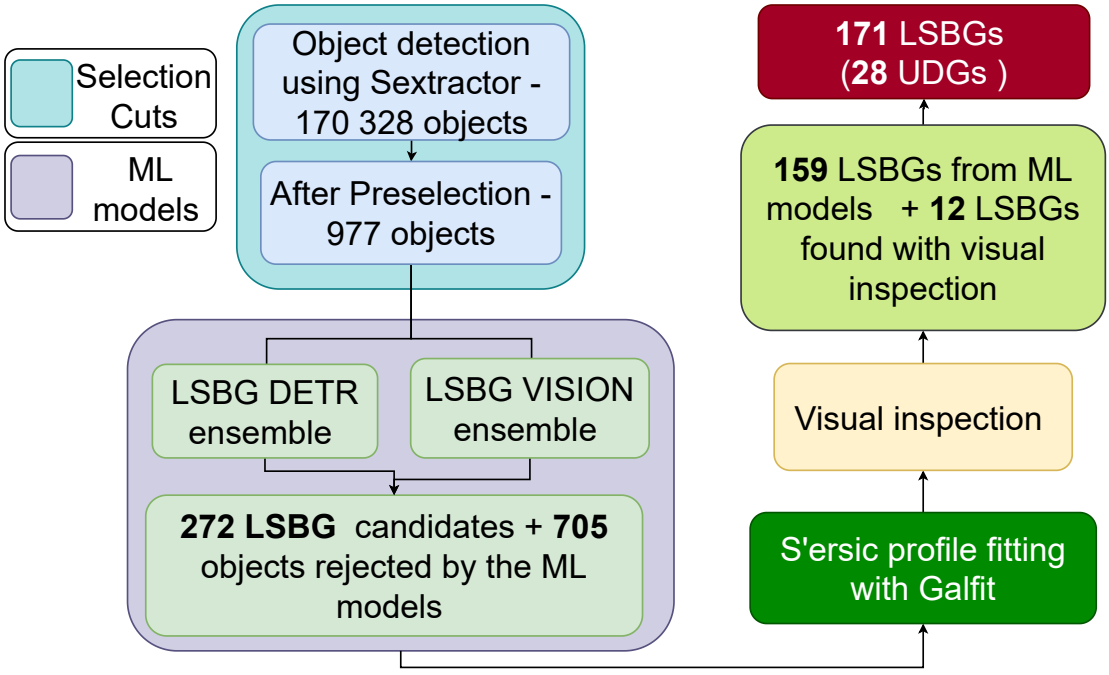
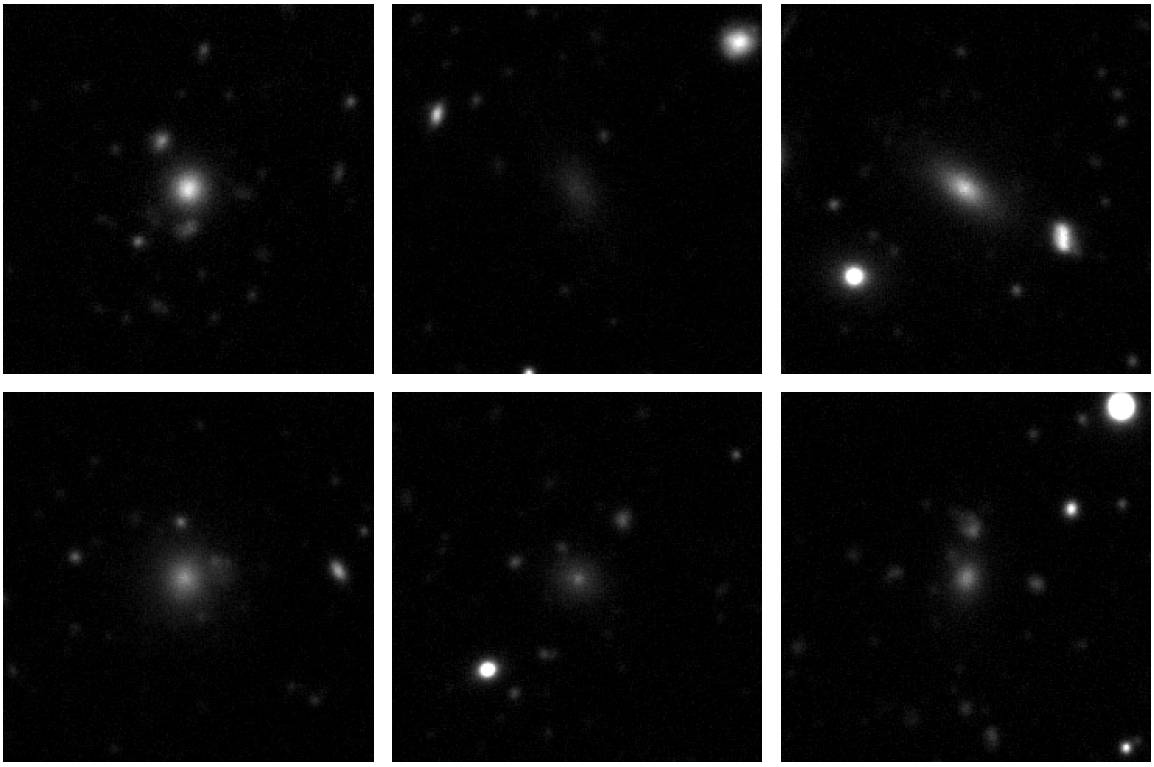


FIGURE 5.6: Schematic diagram showing the sequential selection steps used to find the LSBGs and UDGs in Abell 194.

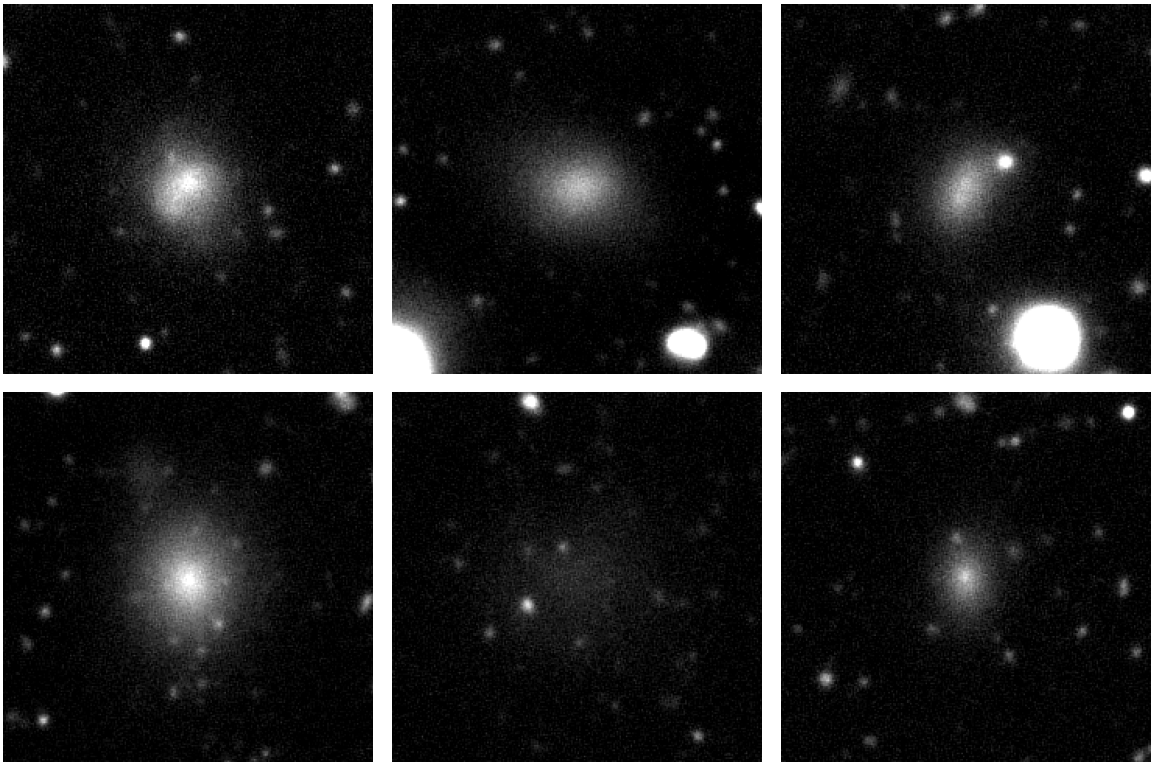
LSBGs are located at distances slightly greater than the virial radius of the cluster where the most distant LSBG is located 1.15 Mpc away from the cluster centre. However, these 12 galaxies are likely to be part of the Abell 194 cluster and here we treat all these 171 LSBGs as part of the Abell 194 cluster. Utilizing the redshift of Abell 194, we determine that 28 LSBGs have a half-light radius greater than 1.5 kpc and a central surface brightness greater than 24.0 mag arcsec^{-2}, thereby classifying them as UDGs. It should be noted that the 28 UDGs are a subset of the 171 LSBs identified in this work.

The distribution of the g-band magnitude of LSBGs and UDGs are plotted in Fig. 5.8a. The UDGs have a median g-band magnitude of 20.73 *mag* which is brighter than the median value of the LSBGs which is 20.23 *mag*, indicating that our sample has very faint LSBGs but are not large enough to be considered as UDGs. The distribution of the Sérsic index (n) of LSBGs and UDGs are plotted in Fig. 5.8b. The Sérsic index of the LSBGs predominantly lies between 0.5 and 1.5 and has a median value of 0.86. However, the UDGs have a Sérsic index distribution with a median of 0.76, which is comparatively lower than the LSBGs indicating that the light distribution of UDGs is more concentrated than that of the LSBGs. A detailed discussion of the properties of the LSBGs and UDGs identified in this work is presented in Sect. 5.5.

As mentioned, in Sect. 5.3.9, a more reliable estimate of the apartment magnitude of the galaxies could be measured from the aperture photometry. We estimated the apparent magnitude in the g,r, NUV and FUV bands along with correcting for galactic extinction as described in Sect. 5.3.9. 20 LSBGs in our sample had confident detections in NUV ($S/N > 3\sigma$), whereas only 15 LSBGs had confident detection in the FUV. For the rest of the sample, we estimated the upper limits of the NUV and FUV magnitude based on their 3σ values within the aperture. The magnitudes measured from the apertures were corrected for Galactic extinction and were used to $g - r$, $NUV - r$ and $FUV - NUV$ colours.

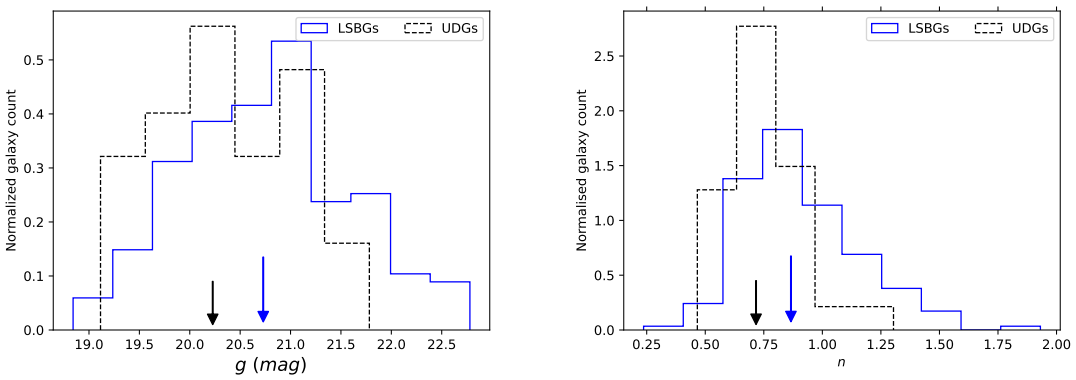


(A) Six examples of LSBGs identified from Abell 194.



(B) Six examples of UDGs identified from Abell 194.

FIGURE 5.7: The top panel (5.7a) shows six examples of LSBGs identified from our study are shown in this work. The bottom panel (5.7b) shows six examples of UDGs identified from our study are shown in this work. Each cutout of the LSBG and UDG corresponds to a $40'' \times 40''$ (152×152 pixel^2) region of the sky centred around the LSBG or the UDG in the g-band.



(A) (B)

FIGURE 5.8: The normalised distribution of the g-band magnitude and Sérsic index for the LSBGs and UDGs presented in this work are shown in the left (5.8a) and right (5.8b) panel respectively. The black arrow shows the median of the LSBG distribution and the blue arrow shows the median of the UDG distribution in both plots.

In addition, the aperture magnitude of the *r*-band was used to estimate the absolute magnitude (M_r) in the *r*-band. Following the stellar mass-to-light ratio vs color relation for LSBGs from Du et al. 2020, we used the g-r color of our sample to estimate the stellar masses ($M_\star$) of the LSBGs. Similarly, the stellar mass surface density (Σ_{star}) of the LSBG sample was calculated using the stellar mass-to-light ratio and the surface brightness in *r*-band following Eq. 1 of Chamba et al. (2022). A detailed discussion on the physical properties of the LSBGs and UDGs found in this work is presented in Sect. 5.5.

5.5 Discussion

5.5.1 Transformers as LSBG Detectors

In this study, we investigate the effectiveness of transfer learning for the identification of LSBGs from datasets deeper than those the model was initially trained on. We have created two different types of transformer ensembles: the LSBG ViT and LSBG DETR ensembles, each comprising four transformer models. These ensembles have been trained on the sample of LSBGS and artefacts identified by Tanoglidis et al. (2021b) and which was further refined by Thuruthipilly et al. (2024b).

On the test datasets from the DES DR1, the ensemble modes achieved an accuracy of 95.62% and a TPR of 96%. In terms of performance, even though there are slight variations in the metrics reported by individual transformer models, both LSBG DETR and LSBG ViT ensemble have almost the same accuracy, TPR, FPR and slightly different AUROC as shown in Table 5.2. Even though both ensemble models have the almost same performance metrics, the distribution of the prediction probabilities of the transformer ensemble is different as seen from Fig. 5.4. In addition, the total number of LSBGs identified by the LSBG DETR and LSBG ViT ensembles independently is also different. For instance, among 171 LSBGs identified in this study, 11 LSBGs were mislabelled as an artefact by either one of the ensemble models.

Another pressing question that demands our attention is why the transformer ensemble models fall short in identifying 12 LSBGs. Even though this accounts for a small fraction of the total number of LSBGs in the data ($\sim$ 8%), we need to have a closer look at

these 12 missed LSBGs to improve the model performance in the future. The magnitude distribution of the LSBGs that has been used to train the ensemble models, LSBGs identified in this work and the LSBGs missed by the model are presented in Fig. 5.9. Here, it should be noted that the magnitude distribution of the training sample used to create the ensemble models and the LSBGs identified by the ensemble models have the same lower limit in magnitude ($g \sim 22.5$) and the number of LSBGs with $21.5 < g < 22.5$ is very low ($\sim 1.7\%$). Consequently, 9 out of the 12 missed LSBGs had magnitudes greater than 21.5 indicating a lack of representation of such faint LSBGs in the training dataset. As a result, the performance of the ensemble models in identifying LSBGs with $g \geq 21$ is suboptimal. Nevertheless, potentially one could refine the ensemble models by retraining them with very faint LSBGs ($g > 22$) from simulations or small-scale deep surveys, thereby extending the models' performance to a fainter regime.

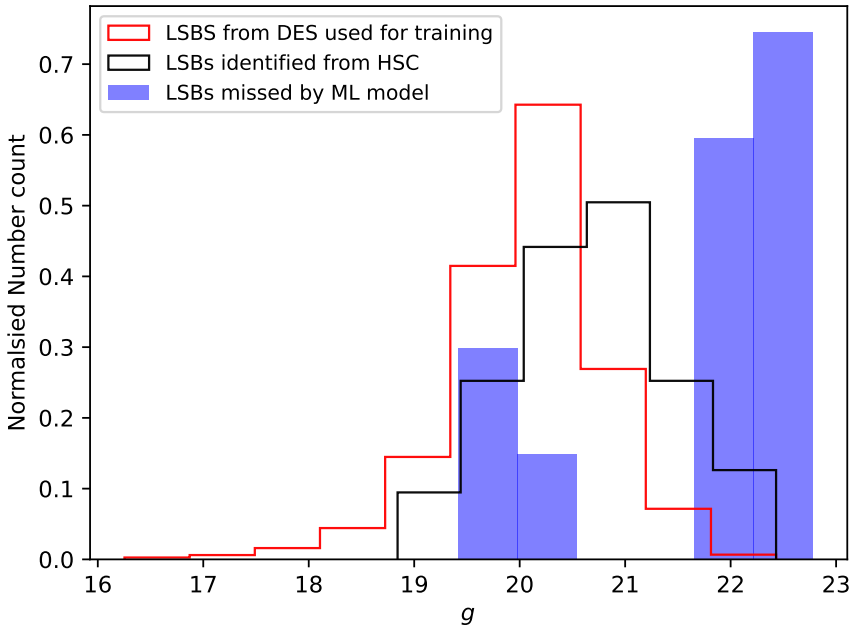
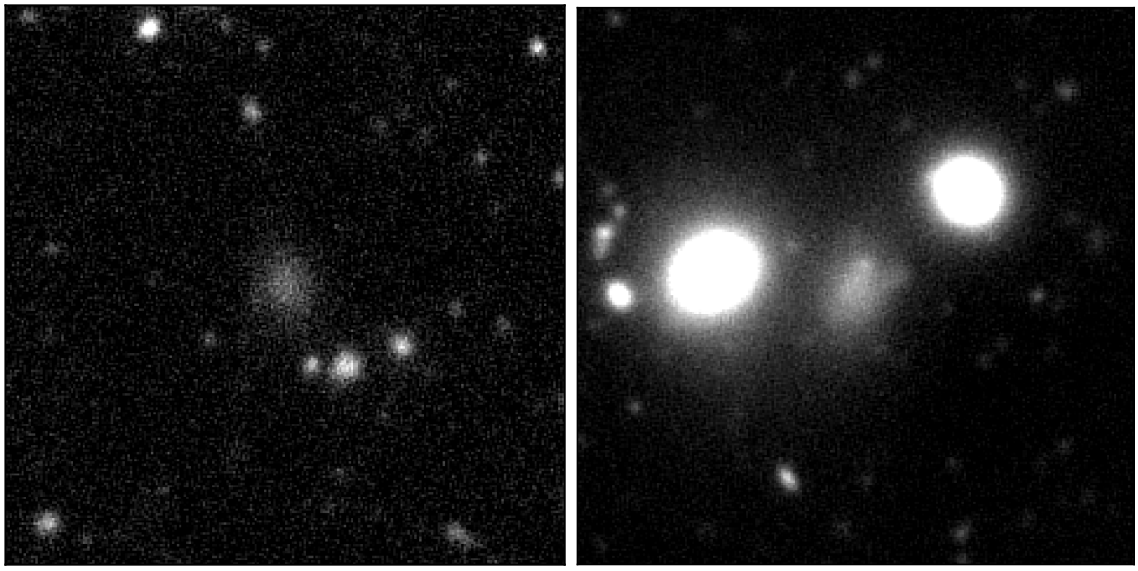


FIGURE 5.9: Normalized distribution of g-band magnitudes for the LSBGs used to train the ensemble models, LSBGs identified by the ensemble models in the HSC data of the Abell 194 cluster, and LSBGs missed by the ensemble models in the HSC data of the Abell 194 cluster.

Similarly, analysing the 3 LSBGs missed by the ensemble model which had ($g < 21$), it was found these galaxies were very close to a bright galaxy which caused the model to classify it as a non-LSBG. Examples of the LSBGs missed by the model are shown in Fig. 5.10. The presence of very bright objects introduces a bias in the prediction probabilities of the ensemble models and it was also found that the effect is more if the bright object is close to the center of the cutout.

Looking at the FPR predicted by the model, even though 101 objects selected as LSBGs by the model were classified as non-LSBGs after galfit and visual inspection, calculating the FPR based on these numbers might be misleading. The rejected sample majorly consisted of three types of objects; galaxies that do not satisfy our LSBG definition but are still faint with $\bar{\mu}_{eff} \in$ 24.0-24.2 mag arcsec^{-2} (around $\sim$ 60), very faint blended objects which are artefacts (around $\sim$ 10) and instrumental artefacts which the model did not have seen during training ((around $\sim$ 30). Simple selection cuts on the morphological parameters



(A) (B)

FIGURE 5.10: Example g-band images of LSBGs missed by the ensemble models. The left panel shows an LSBG which is fainter than the training sample and the right panel shows an LSBG very near to bright galaxies.

can remove the first class. Similarly, the third class of objects could be removed by having a representative sample of these types of instrumental artefacts in the training set or applying selection cuts on the goodness of a Sérsic fit since most of the instrumental artefacts will not converge to give a good Sérsic fit. The second class of contaminants are generally very faint and are hard to remove and need a visual inspection, but their numbers are very low compared to the total number of LSBGs. It should be noted that even human inspectors will have trouble finding very faint sources. For instance, 17 sources for which the two authors disagreed during visual inspection had 14 sources with $\bar{\mu}_{eff} > 26.0$ mag arcsec^{-2}.

One additional point to be considered from the perspective of LSST or Euclid is the load of fine-tuning the masks for LSBGs. We are expected to find more than 100 000 LSBGs in these surveys and ensuring the quality of the masks used for Sérsic profile fitting with visual inspection would be a time-consuming task. In this scenario, one possible solution is the development of DL-based segmentation models that could efficiently mask the contaminants and surroundings of a galaxy. Similarly, citizen science projects could be used to confirm the predictions of a DL model for large-scale surveys.

5.5.2 Previous catalogues of LSBGs and UDGs in Abell 194

In total, Tanoglidis et al. (2021b) and Thuruthipilly et al. (2024b) identified 96 LSBGs in Abell 194, which we call 'LSBGs in DES' for simplicity from here onwards. Among the LSBGs in DES, there were 19 LSBGs which could be classified as UDGs based on their physical size and central surface brightness in the g-band. Similarly, Zaritsky et al. (2022) have identified 14 UDGs in Abell 194 among which 9 have been included in the 'LSBGs in DES' sample.

In our final sample, 13 of the UDGs from Zaritsky et al. (2022) were identified by the ensemble model and one UDG was later identified when searching through the rejected samples. Similarly, 95 of 96 LSBGs from the DES sample have been re-identified by the ensemble model and one LSBG classified as non-LSBG by the model was later found during visual inspection. However, in our final catalogue, there are only 84 LSBGs from the

LSBGs in the DES sample, the remaining 12 have been classified as non-LSBGs since 11 galaxies were found to be slightly brighter (24.0 $<\bar{\mu}_{eff_g} <$ 24.2) than our definition and one LSBG was found to be at a higher redshift (z=0.15) when crossmatched with the SDSS DR 16 sample. Similarly, 7 LSBGs classified as UDGs based on the measurements from DES were also rejected as non-UDGs because 6 of them had a physical radius of less than 1.5 kpc and one of them had $\mu_{0,g}$ less than 24.0 mag arcsec^{-2}.

One of the main reasons for the rejection of some LSBGs and UDGs from the previous catalogues is because of the improved masking strategy that we used in this work that removed small contaminants on top of the galaxies as shown in Fig. 5.3. In shallow data, these contaminants might not be clearly visible, but with deeper data, these contaminants are clearly visible and significantly affect the estimation of the morphological parameters such as Sérsic index and half-light radius during a Sérsic fit. A comparison of the Sérsic index measurement of the 84 LSBGs identified commonly from LSBGs in DES and from this work with HSC is shown in Fig. 5.11. The measurements from DES were performed without masking the contaminants that may reside on top of the LSBGs. As these contaminants are very compact objects, the presence of these contaminants tends to increase the Sérsic index of the LSBG during fitting which is clear from Fig. 5.11. Similarly, because of the additional light from these contaminants the $r_{1/2}$ of these galaxies also tends to be over-estimated. However, the presence of these contaminants does not significantly affect the magnitude of the galaxy. These trends were also observed for the parameter estimation of the UDGs in the Coma cluster by Bautista et al. (2023) who also used similar masks to ours.

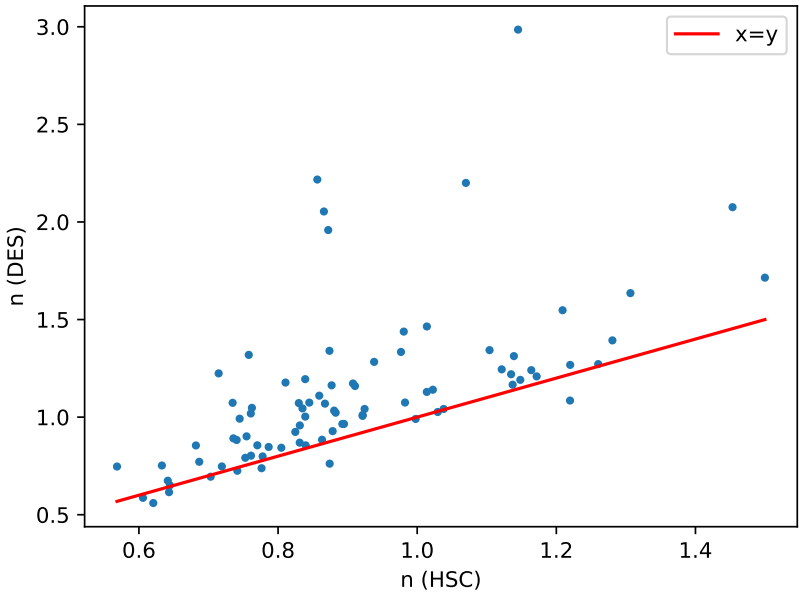


FIGURE 5.11: Comparison of Sérsic index measurement from DES and HSC. The y-axis shows the Sérsic index of LSBGs estimated without masking the contaminants reported in Tanoglidis et al. (2021b) and Thuruthipilly et al. (2024b) from the DES data. The x-axis shows the Sérsic index of the same LSBGs estimated by masking the contaminants in this work.

Furthermore, an additional factor contributing to the deviation in the Sérsic parameters between this study and those reported by DES is the local sky subtraction. Previous searches for LSBGs in DES by Tanoglidis et al. (2021b) and Thuruthipilly et al. (2024b) did

not correctly consider the local sky subtraction. The publicly available DES data have estimated the sky background with regions of size $\sim 1'$ (Morganson et al. 2018; Bernstein et al. 2018) which is considerably larger than the size of the LSBGs we investigate in this work. Sky subtraction over a large area tends to over-subtract the sky influenced by the presence of bright objects in the data. Which sometimes results in the LSBGs appearing fainter than their true surface brightness. Failing to correct for local sky subtraction can lead to inaccurate Sérsic fits, as also highlighted by Bautista et al. (2023).

Despite the exclusion of some LSBGs from the previous DES sample, the number of LSBGs in the Abell 194 cluster has doubled from 84 to 171. This significant increase is partly attributed to the deeper sensitivity of the HSC data, exceeding that of the DES data by an order of magnitude. However, a close inspection of the LSBG sample shows that the increase in numbers is not solely due to increased depth. The preselection criteria applied by `Sextractor` also contribute substantially. For instance, the ratio of $r_{1/2}$ estimated with `Sextractor` and `Galfit` for LSBGs presented in this work has a median value of 0.8 suggesting that $r_{1/2}$ estimated by `Sextractor` yields a slightly smaller radius than from `Galfit`. A similar value (~ 0.8) for the ratio of $r_{1/2}$ estimated with `Sextractor` and `Galfit` is found for the LSBG sample presented by Tanoglidis et al. (2021b). Hence, our preselection criteria were relaxed to `FLUX_RADIUS_G` $> 2.0''$ to account for the selection bias from the `Sextractor` in comparison to the preselection condition of `FLUX_RADIUS_G` $> 2.5''$ used by Tanoglidis et al. (2021b) and Thuruthipilly et al. (2024b). Consequently, 62 out of the 87 newly identified LSBGs have `FLUX_RADIUS_G` $> 2.0''$. With the same preselections as in Tanoglidis et al. (2021b) or Thuruthipilly et al. (2024b), there are only 25 new LSBGs which is an increment of $\sim$ 30%. In the case of the number of UDGs, the preselection does not significantly have an impact as only 3 UDGs out of 28 have `FLUX_RADIUS_G` in the range $2 - 2.5''$ since 1.5 kpc corresponds to $4.12''$ at the redshift of Abell 194 cluster. However, if the cluster was situated slightly further away ($z \gtrsim 0.025$) compared to Abell 194 ($z = 0.017$), then the effects of preselection become significant as the angular size of a UDG will be comparable to $2.5''$.

The radar plot comparing the properties of the LSBGs identified from DES and LSBGs identified with HSC in Abell 194 is shown in Fig. 5.12. In terms of color $(g - r)$, q and n, both samples have the same median indicating that both samples are similar in these properties. As mentioned earlier, the new sample of LSBGs has many small LSBGs in size which is evident from the median values of the LSBGs detected from DES (1.43 kpc) and not detected in DES (1.13 kpc). Similarly, as expected the new sample of LSBGs is fainter than the LSBGs detected in DES which is seen from the high median values of $\bar{\mu}_r$, $\mu_{0,g}$, $\log\Sigma_\star$, $\log M_\star$ and $r_{aperture}$ of LSBGs detected in DES compared to the LSBGs not detected in DES.

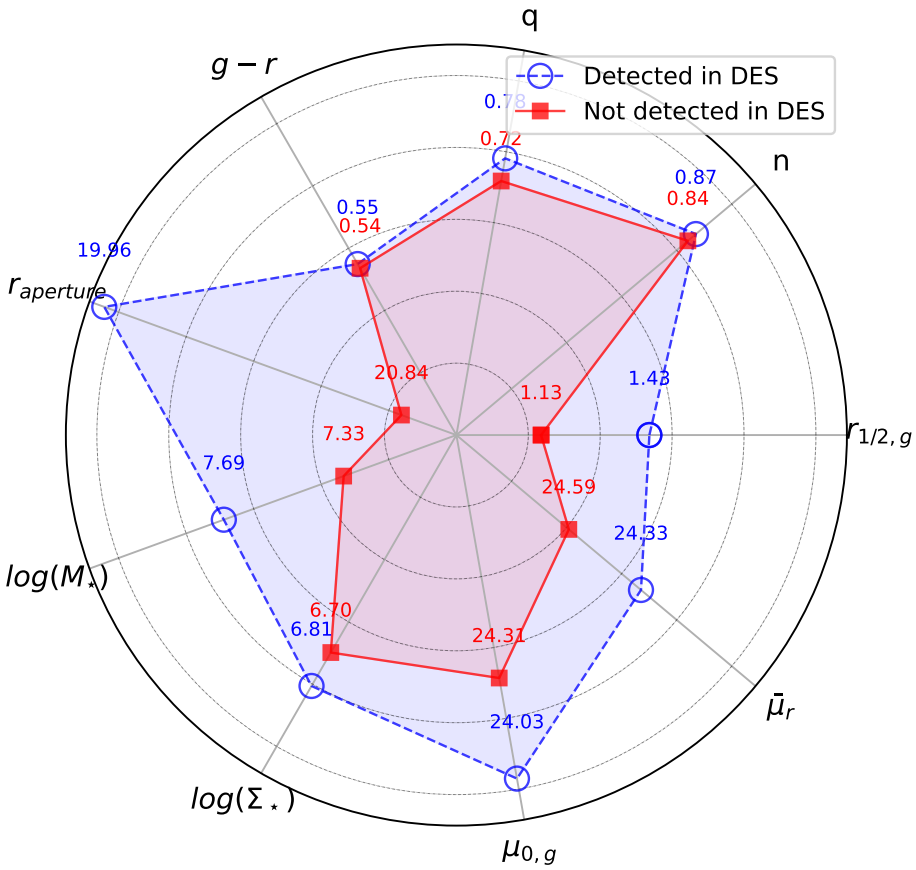


FIGURE 5.12: Comparison of morphological and physical properties of LSBGs identified from DES and the LSBGs not identified from DES. The median of $g-r$, q, n, $r_{eff,g}$ (kpc), $\bar{\mu}_r$ (mag arcsec^{-2}),$\mu_{0,g}$ (mag arcsec^{-2}), log$\Sigma_\star$ ($M_\odot$), log$M_\star$ ($M_\odot$ kpc^{-2}) and $r_{aperture}$ are shown in the cyclic order.

5.5.3 LSBG and UDG population of Abell 194 cluster

The catalogue of the LSBGs identified in Abell 194 along with its estimated properties are given in Table. 5.3. Combining transformers along with visual inspection, in total we find 171 LSBGs. As mentioned in Sect. 5.1, UDGs are a subclass within LSBGs, characterized by their extended half-light radii ($r_e \geq 1.5$ kpc) and high central surface brightness ($\mu_0 > 24$ mag/arcsec^2) in the g-band (van Dokkum et al. 2015a). Assuming all the LSBGs found in this work share the same redshift as Abell 194, we estimate that our sample has 28 UDGs and 143 non-UDGs. It should be noted some of these galaxies may be foreground or background galaxies and the individual redshifts of these galaxies are needed to confirm the status of being a UDG.

Since van Dokkum et al. (2015a) found a substantial population of UDGs within the Coma cluster, the subsequent studies by Koda et al. (2015); Yagi et al. (2016); Mihos et al. (2015); Lim et al. (2020); La Marca et al. (2022b); Bautista et al. (2023) have identified numerous UDGs in various galaxy clusters. Further investigations have demonstrated that not only cluster, groups are also abundant with UDGs (Merritt et al. 2016; Román & Trujillo 2017a; Müller et al. 2018; Somalwar et al. 2020; Forbes et al. 2020) and their abundance has

a near-linear correlation with the mass of the dark matter halo of the cluster (van der Burg et al. 2017; Mancera Piña et al. 2018; Karunakaran & Zaritsky 2023). The number of UDGs as a function of host halo mass including the Abell 194 is shown in Fig. 5.13. Karunakaran & Zaritsky (2023) estimated that the number of UDGs scale as a function of the halo mass following the relation $N = (38 \pm 5)(\frac{M_{200}}{10^{14}})^{0.89\pm0.04}$. For the Abell 194 cluster which has a halo mass of $7.6 \times 10^{13} M\odot$, this relation predicts 30 ± 4 UDGs.

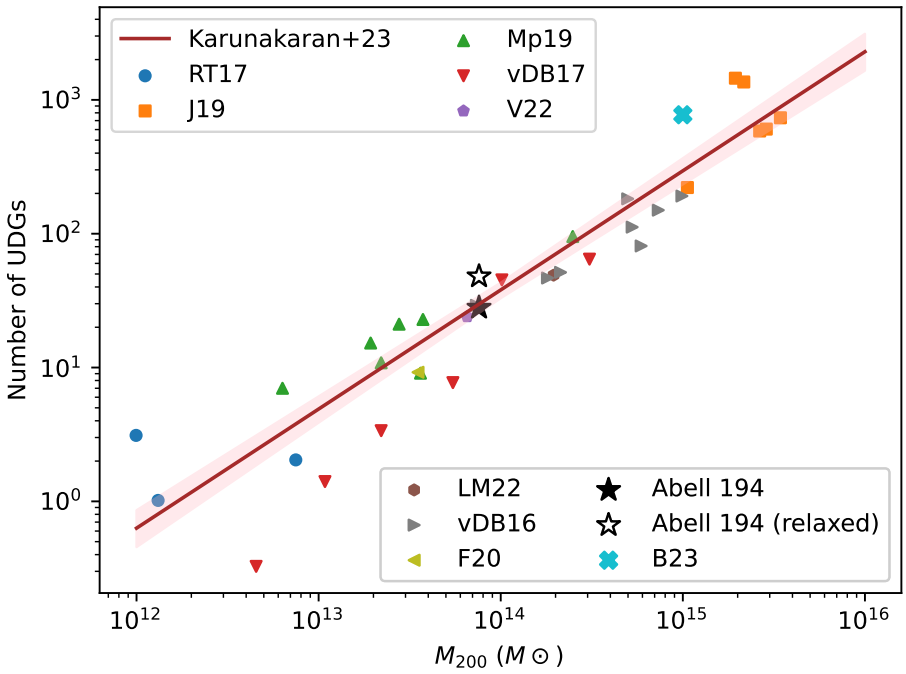


FIGURE 5.13: Number of UDGs as a function of halo mass. The cluster UDGs are from van der Burg et al. (2016); Janssens et al. (2019); Mancera Piña et al. (2019); Venhola et al. (2022); La Marca et al. (2022a); Bautista et al. (2023) and are represented by and group UDGs are from van der Burg et al. (2016), Román & Trujillo (2017a), Forbes et al. (2020)

However, since the UDG definition in the literature is not unique, the sample used by Karunakaran & Zaritsky (2023) is brighter ($\bar{\mu_g} > 24.0$ mag arcsec^{-2}) than our UDG sample which has $\mu_{0,g} > 24.0$ mag arcsec^{-2}. Using the same definition as used by Karunakaran & Zaritsky (2023) we have 45 LSBGs satisfying the criteria to be a UDG. This is greater than the value predicted by Karunakaran & Zaritsky (2023) and shown in Fig. 5.13. As our primary objective does not involve constraining the relationship between halo mass and the number of UDGs, we have refrained from recalibrating this relation and have reserved it for potential future extensions of this study.

As the UDGs are considered a subclass of LSBGs, one might be curious about the difference in the properties of UDGs and non-UDGs (LSBGs that did not satisfy the condition to be a UDG). The radar plot comparing the median of the distribution of all the properties estimated in this work for the UDGs and non-UDGs is shown in Fig. 5.14. On comparing the median value of $g - r$ color, we can see that both UDGs and non-UDGs have similar color distribution. The median color of the UDGs reported here is 0.53 which is similar to the median color of the UDG sample presented in Venhola et al. (2022); Mancera Piña et al. (2019) and Bautista et al. (2023). Similarly, in terms of axis ratio also both the UDGs and

non-UDGs have similar distribution and have a median value of $\sim$ 0.75 indicating that LSBGs in this sample have a slightly elongated shape.

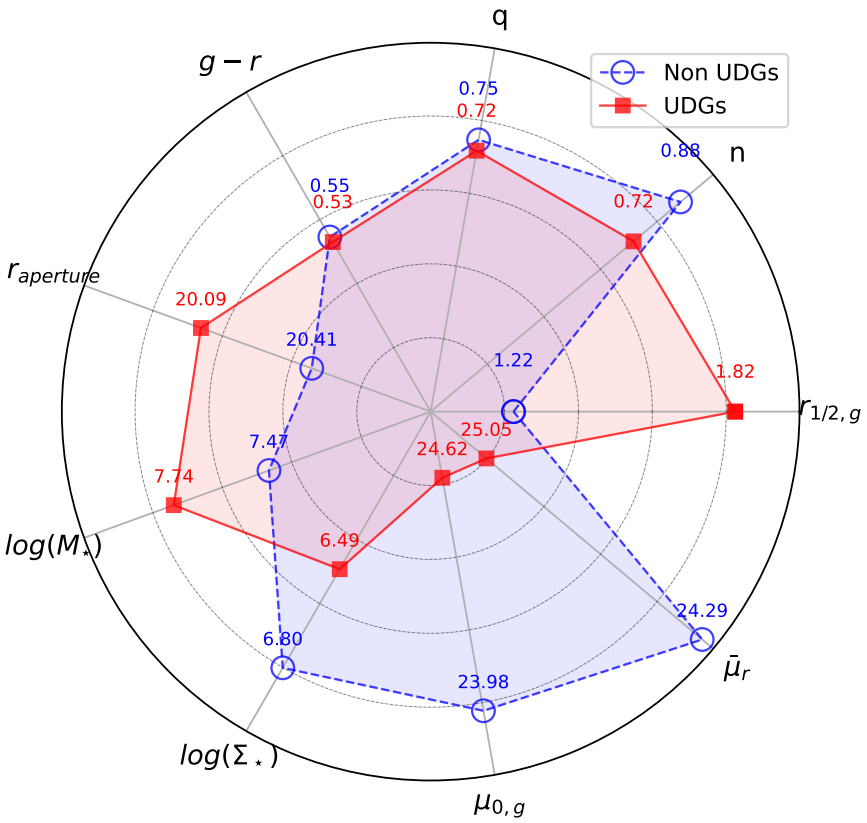


FIGURE 5.14: Comparison of morphological and physical properties of UDGs and non-UDGs. The median of $g-r$, q, n, $r_{eff,g}$ (kpc), $\bar{\mu}_r$ (mag arcsec^{-2}), $\mu_{0,g}$ (mag arcsec^{-2}), log$\Sigma_\star$ ($M_\odot$), log$M_\star$ ($M_\odot$ kpc^{-2}) and $r_{aperture}$ are shown in the cyclic order.

Coming to the Sérsic index, the non-UDGs ($n \sim 0.88$) have slightly higher median value compared to UDGs ($n \sim 0.72$). These values are consistent with the values reported for the UDGs in Venhola et al. (2022); Mancera Piña et al. (2019) and Bautista et al. (2023). In addition, it should be pointed out that the Sérsic index distribution of non-UDGs and UDGs are similar to the case of dwarf ellipticals as observed by Poulain et al. (2021), suggesting that a significant portion of the sample could be comprised of such sources.

The median value of half-light radius in the g-band of UDGs (1.83 kpc) is higher than non-UDGs (1.22 kpc) is an effect of UDG being defined as extended sources. Similarly, The median values of central surface brightness in the g-band (UDG: 24.62 mag arcsec^{-2}, non-UDG: 23.98 mag arcsec^{-2}), mean surface brightness in the $r-$ band (UDG: 25.05 mag arcsec^{-2}, non-UDG: 24.29 mag arcsec^{-2}) and stellar mass surface density (UDG: $10^{6.8}$ M$_\odot$ kpc^{-2}, non-UDG: $10^{6.49}$ M$_\odot$ kpc^{-2}) are higher for UDGs by the effect of the definition of UDGs. However, it should be noted that the median of the total stellar mass and apparent magnitude in the r-band of UDGs are greater than non-UDGs. This indicates that the

UDGs being extended sources have more stellar mass and produce more light in comparison to the non-UDGs.

5.5.4 Cluster-centric properties of LSBGs and UDGs

The spatial distribution of LSBGs and UDGs in the Abell 194 along with the spatial density of the LSBG is presented in Fig. 5.15. The Abell 194 has a tail of bright galaxies from the cluster centre to the direction of NGC 519 which are found to be falling towards the cluster centre (Tempel et al. 2016). The spatial density of the LSBGs at the cluster centre is aligned with the bright galaxies in the direction of NGC 519 potentially, pointing that the LSBGs follow the bright galaxies to the cluster centre and the LSBGs might be the satellite galaxies of the bright galaxies.

The projected distance is often used as a proxy for density within clusters and can be used to analyse how the environment affects the structural properties of an LSBG or UDG. The structural and physical properties of the LSBGs and UDGs as a function of the cluster-centric distance are shown in Fig. 5.16. To facilitate comparison, we assign the area within a radius of less than 0.4 Mpc around the cluster centre as the inner region of the cluster, while the region beyond a radius of 0.8 Mpc is considered the outer skirts of the cluster. The region in between is considered the middle region of the cluster.

The median value of the Sérsic index exhibits a small increment ($\sim$ 0.1) for LSBGs and UDGs in the inner regions of the cluster compared to its outskirts. However, overall, there is no significant trend observed in the Sérsic index of both LSBGs and UDGs as a function of the cluster-centric distance. Meanwhile, the median of the axis ratio of the LSBGs tends to be higher near the cluster centre ($b/a \sim 0.8$) compared to the outskirts of the cluster ($b/a \sim 0.7$). Similarly, the median values of UDGs tend to be $\sim b/a \sim 0.7$ near the inner region of the cluster and decrease to $b/a \sim 0.6$ in the outer region of the cluster (>0.8 Mpc). For both LSBGs and UDGs, we notice a slight bump at a distance of 0.8 Mpc from the cluster centre. Our observations on the trends in the Sérsic index and axis ratio, align well with observation with Mancera Piña et al. (2019) and are in good agreement with expectations from models of dwarf galaxies that have undergone harassment and tidal interaction processes (Moore et al. 1996; Aguerri & González-García 2009)

The surface brightness distribution presented in Fig. 5.16 shows that LSBGs and UDGs in the inner regions of the cluster and they are brighter in comparison to the middle region of the cluster. The lack of faint LSBGs and UDGs in the inner regions might be because of the destruction of low-mass galaxies by the highly massive galaxies in the cluster centre. Another possible reason for this could be the non-detection of the faint sources due to the contamination from the bright sources and the intra-cluster light. However, it should be also noted that the outer region of the cluster shows a lack of very faint LSBGs. A detailed follow-up study with improved methodology is necessary to confirm if this is a statistical bias or corresponds to any physical phenomena.

Recently, Román & Trujillo (2017a) reported a decrease in the effective radius of UDGs as they lie closer to the cluster centre. In contrast, we have some large UDGs in the inner regions of the cluster centre. This discrepancy might be due to a projection effect. Alternatively, it could be also due to that the Abell 194 cluster has a different trend compared to the clusters studied by Román & Trujillo (2017a). As for LSBGs, we do not see any clear trends and the running median is relatively flat. The trend in the stellar mass distribution of our sample and the trend reported by Román & Trujillo (2017a) are also not in agreement with each other. Román & Trujillo (2017a) reported a decrease in the stellar mass as we go towards the cluster centre whereas for our sample we find that the stellar mass in the inner region of the cluster is relatively high in comparison to the stellar mass of galaxies in the outskirts.

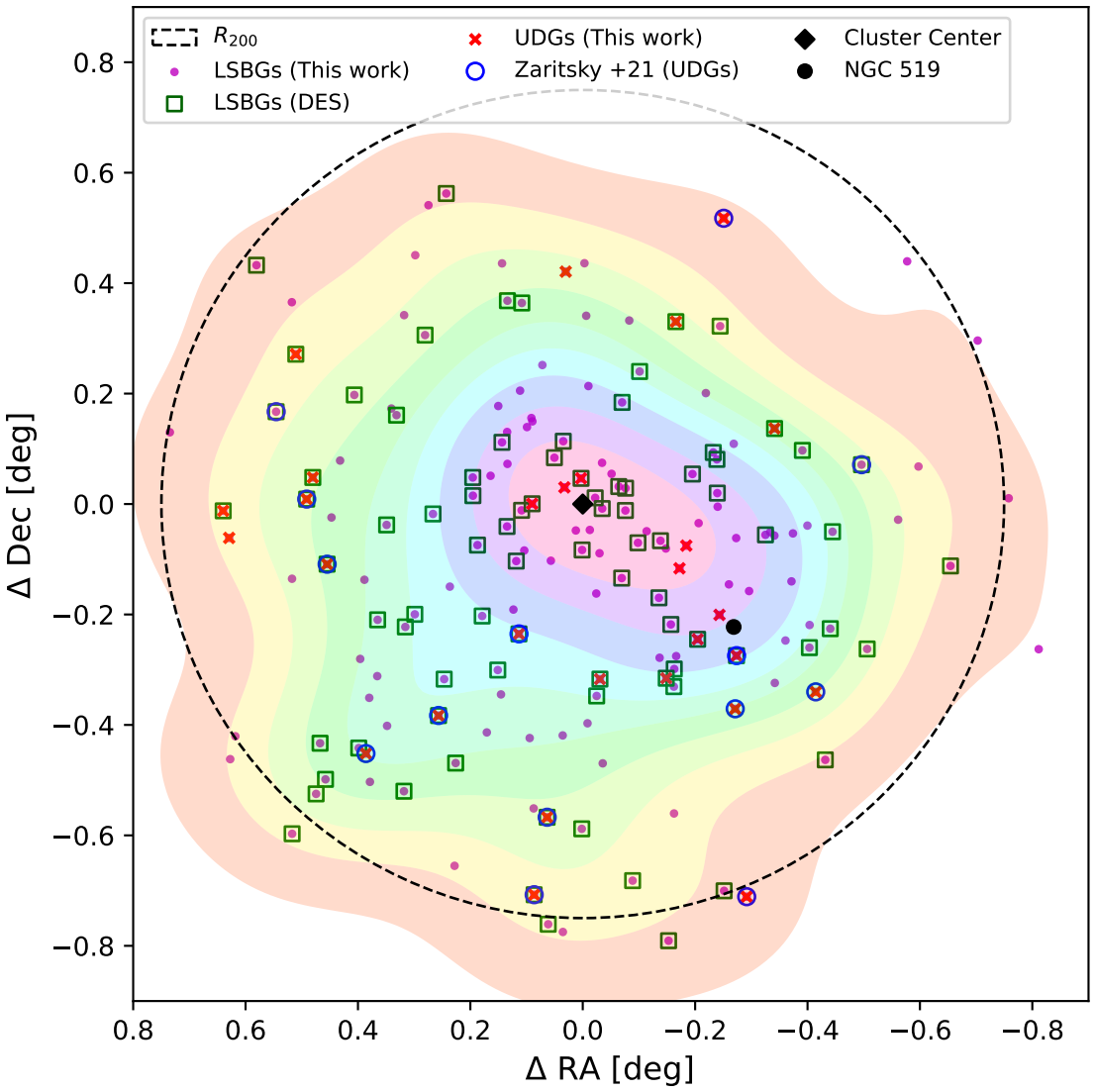


FIGURE 5.15: Spatial distribution of the LSBGs and UDGs identified in this work along with the known LSBGs and UDGs around the Abell 194. The spatial density of the LSBGs in Abell 194 is presented as contours. The LSBGs and UDGs identified in this work are shown in magenta dots and red upward triangles, respectively. The green rectangles represent the combined LSBG sample from Tanoglidis et al. (2021b) and Thuruthipilly et al. (2022). The blue circle shows the UDGs reported in Zaritsky et al. (2022). The black dashed circle, the black diamond and the black dot represent the virial radius, cluster centre and the NGC 519.

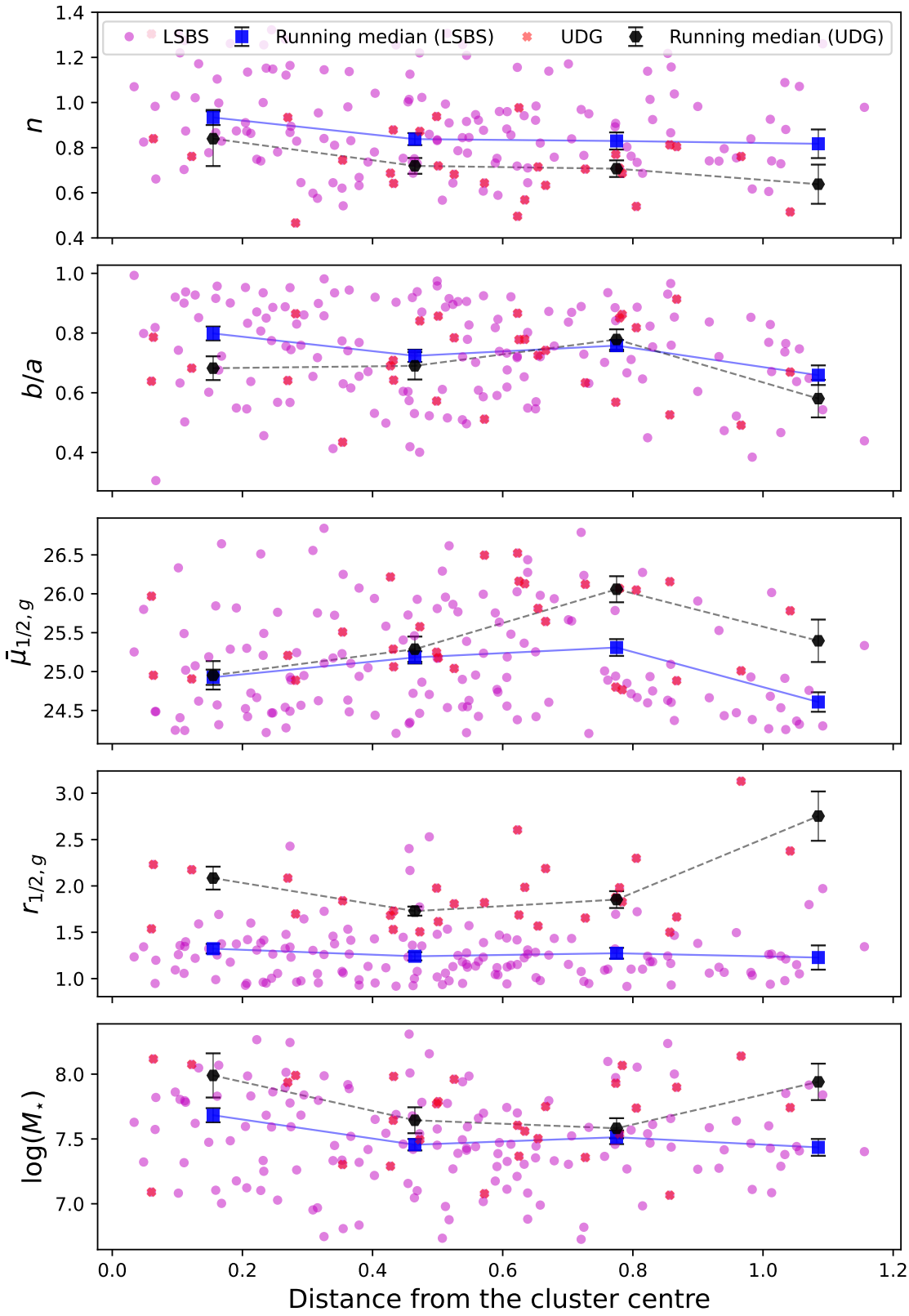


FIGURE 5.16: Morphological and physical properties of LSBS and UDGs as a function of cluster centric distance. The LSBGs and UDGs identified in this work are shown in magenta dots and red crosses, respectively. The running median of LSBGs and UDGs are calculated with a bin size of 0.33 Mpc and are represented as blue and black dotted lines respectively.

Recently, Venhola et al. (2019b) found that the early type dwarf galaxy population in the Fornax cluster becomes slightly redder as we go towards the cluster centre in the $u - X$ where $(X \in g, r, i)$ color. UDGs also have shown a similar trend in becoming bluer as going towards the outskirts of the cluster centre (Román & Trujillo 2017a; Mancera Piña et al. 2019; Junais et al. 2022). The colours $FUV - NUV$, $NUV - r$ and $g - r$ as a function of the cluster centric distance are shown in Fig. 5.17. Similar to the observations of Venhola et al. (2019b) and Junais et al. (2022) we can see that galaxies become redder in $FUV - NUV$ and $NUV - r$ colours as we go towards the the cluster centre. These colors ($FUV - NUV$ and $NUV - r$) are more sensitive towards the recent star formation rate. Hence, the visible trend in these colours indicates that the LSBGs in the outskirts have more recent star formation than those LSBGs near the cluster centre. On the other hand, our sample of LSBGs does not show any trend in $g - r$ as a function of the cluster-centric distance up to 0.8 Mpc. However, the LSBGs in the outer parts of the cluster become bluer in comparison to the LSBGs in the inner and middle regions of the cluster. Similarly, the LSBGs that have NUV detection in the inner region of the cluster are redder in the $g - r$ color compared to the LSBGs in the middle and outer regions of the cluster.

The similarities in the trends of cluster-centric properties and morphological properties for LSBGs and dwarf galaxies indicate that the majority of our LSBG population is composed of dwarf galaxies. Recently, Lazar et al. (2024) have found that the red to blue fraction of the dwarf galaxies have similar ratios as LSBGs supporting the argument that the majority of LSBGs are indeed dwarf galaxies. Hence, these results align perfectly with the argument in the literature that the UDGs are a subset of dwarf galaxies (Conselice 2018; Benavides et al. 2023).

5.5.5 Distribution in Color–Magnitude Space

Ultraviolet (UV) observations offer valuable insights into the formation of UDGs and LSBGs. As mentioned earlier, UV observations are sensitive to recent star formation, providing clues to the assembly process of galaxies. Recently, Singh et al. (2019) studied the UV properties of UDGs in the Coma cluster using the GALEX data and concluded that UDGs in the cluster are mostly quiescent without any recent star formation. However, a subsequent analysis by Lee et al. (2020) showed that in some UDGs near the cluster centre, there could be recent star formation in the cluster UDGs which is consistent with our findings.

The color-magnitude space can be used to separate the galaxy sample into quiescent red galaxies and blue star-forming galaxies. In Fig, 5.18, we compare the distribution of the LSBGs and UDGs from our sample and the UDGs from the Coma cluster from Singh et al. (2019) and Lee et al. (2020) in the color-magnitude space ($NUV - r$ vs M_r). In our sample, we have only one UDG with NUV detection which belongs to the blue population. Similar to the population of Coma UDGs from Lee et al. (2020), we have two LSBGs belonging to the red population with NUV detection. It should be also noted that the population of LSBGs and UDGs in our sample are much fainter than the population of UDGs presented in Singh et al. (2019) and Lee et al. (2020). Our detections are consistent with the extension of red and blue luminous galaxies from SDSS.

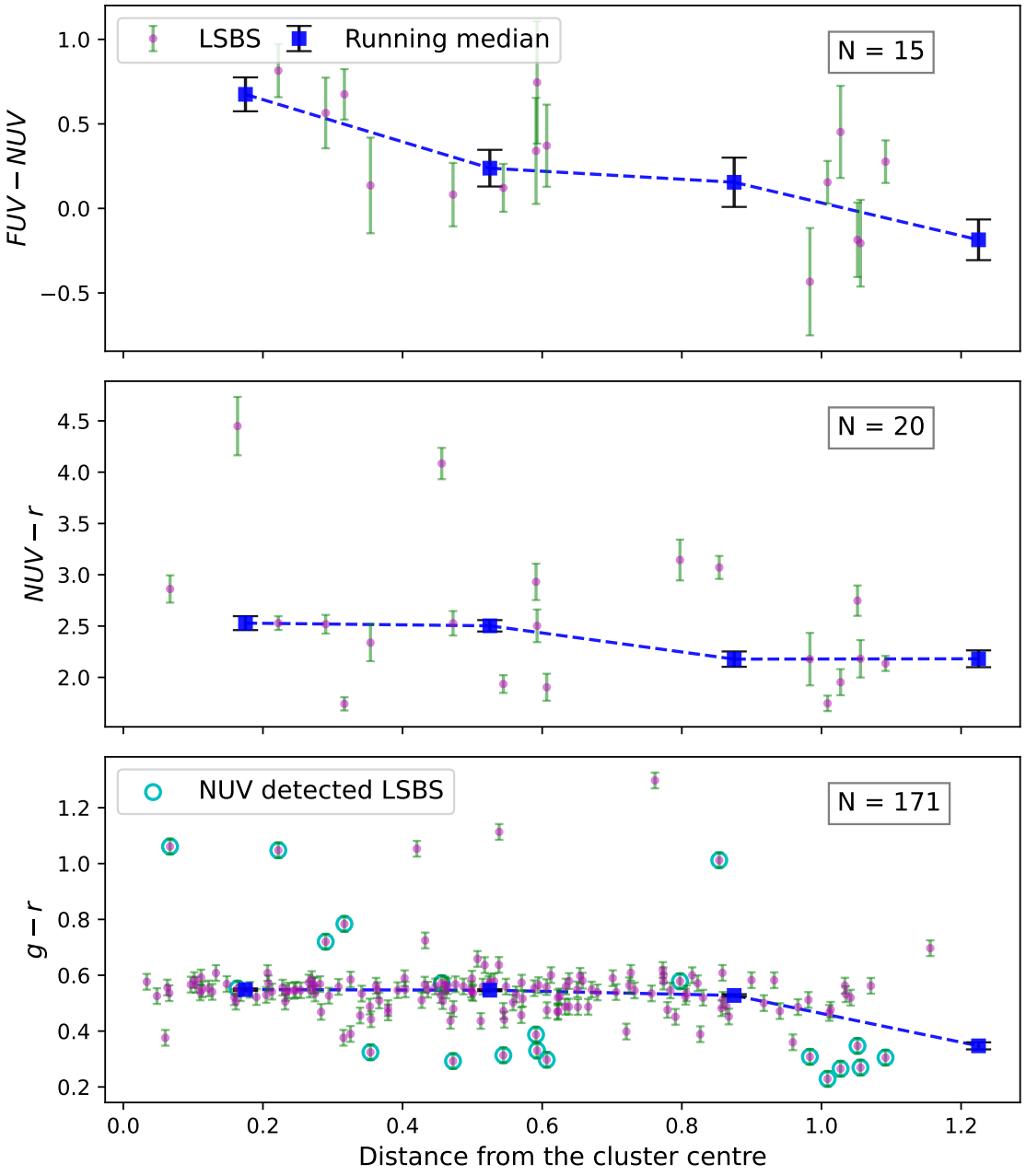


FIGURE 5.17: $FUV - NUV$ (top panel), $NUV - r$ (middle panel) and $g - r$ (bottom panel) colors of the LSBGs presented in this works as a function of cluster centric distance. The number of LSBGs used is also shown in each subplot since not all the LSBGs had 3σ detection in NUV and FUV bands

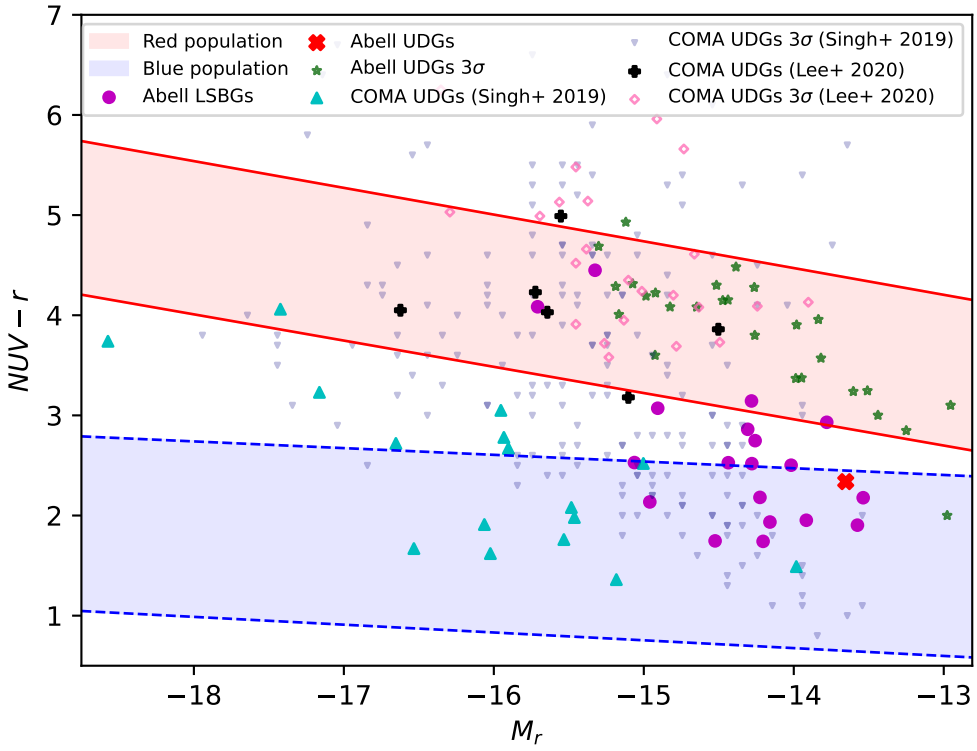


FIGURE 5.18: The $NUV - r$ color is plotted against the absolute magnitude in the r-band. The red region denotes the space occupied by the red-quiescent galaxy population, while the blue region represents the blue galaxy population (Singh et al. 2019). Magenta dots represent LSBGs and red crosses represent UDGs with NUV detection in our sample. Green stars indicate the upper limits of $NUV - r$ for UDGs in Abell 194 without NUV detection. Cyan upward triangles and black plus symbols represent UDGs from the Coma cluster reported by Singh et al. (2019) and Lee et al. (2020), respectively. Downward navy blue triangles and pink rhombuses denote the upper limits of $NUV - r$ for UDGs in the Coma cluster from Singh et al. (2019) and Lee et al. (2020), respectively.

5.6 Conclusion

In this paper, we explore the possibilities of identifying LSBGs from the dedicated observation data of cluster Abell 194 with transfer learning using the transformer ensemble models presented in Thuruthipilly et al. (2024b). We trained the transformer ensemble models on the LSBG and artefacts from DES DR 1 which is presented in Tanoglidis et al. (2021b) and Tanoglidis et al. (2021a) and updated by Thuruthipilly et al. (2024b). Subsequently, the trained transformer models were tested on the data of the Abell 194 cluster images with HSC which is deeper than the DES DR 1 by two orders of magnitude. Since the instrument used to image in DES DR 1 is different from the HSC, we standardise the data from both instruments by converting their units into surface brightness units (μJy arcsec^{-2}). On the standardised DES DR 1 data the transformer ensemble models achieved an accuracy of 95% which is comparable to the $\sim$ 94% accuracy of the ensemble models presented in Thuruthipilly et al. (2024b).

From the HSC data of the cluster Abell 194, we identified 171 LSBGs and 28 UDGs which are a subset of the LSBG population. Applying aperture photometry to the available GALEX data of Abell 194, we discovered that 15 LSBGs in our sample emit radiation in both NUV and FUV wavelengths, while 5 LSBGs emit radiation only in the NUV wavelength, indicating recent star formation in these galaxies.

The trained ensemble models achieved a recovery rate of 93% and correctly identified 159 LSBGs and missed 12 LSBGs. We found that among the 12 LSBGs missed by the ensemble models, 9 LSBGs are fainter than the training sample ($g < 21.5$ mag) and do not have a representative sample in the training set. The rest 3 LSBGs missed have bright objects near the centre of the LSBGs. The reported metrics of the transformer models without fine-tuning confirm the success of transfer learning and these models could be potentially applied to deeper datasets.

In this work, we also found that the improved masks and local sky subtraction can strongly influence the parameters estimated with Sérsic fit which was also noted by Bautista et al. (2023). Among the affected parameters it was found that the Sérsic index of the LSBGs is found to be mostly affected where as the magnitude is the least affected. As a result, we reject 11 LSBGs found in DES DR 1, as the $\bar{\mu}_{eff}$ estimated with HSC using improved masks is slightly brighter (24.0 <24.2 mag arcsec^{-2}) than the surface brightness threshold (24.2 mag arcsec^{-2}).

The newly estimated number of UDGs in Abell 194 with the data from the HSC in this work support the argument in the literature that the number of UDGs in a cluster is linearly correlated with the cluster mass (van der Burg et al. 2017; Mancera Piña et al. 2018; Karunakaran & Zaritsky 2023). However, we also note that the number of UDGs found in Abell 194 is greater than the predicted number based on the relation from Karunakaran & Zaritsky (2023). Since we did not have enough new data points to correct the relation statistically, we left the subsequent analysis in this direction for the future. The S'ersic index distribution of the UDGs reported in this work is comparable to the values reported in the literature (Koda et al. 2015; Yagi et al. 2016; Román & Trujillo 2017a; van der Burg et al. 2017; Venhola et al. 2022; Bautista et al. 2023; Thuruthipilly et al. 2024b). In addition, it should be noted that the S'ersic index distribution of the UDGs is also comparable to that of dwarf elliptical galaxies, suggesting that UDGs might be the tail end of the dwarf galaxy population.

Analyzing the cluster-centric properties of the LSBGs and UDGs in the Abell 194 cluster revealed that those near the cluster centre tend to have higher S'ersic indices than those in the outer regions of the cluster which is consistent with the observations reported in Mancera Piña et al. (2019). The brighter LSBGs and UDGs are also found closer to the cluster centre, however, we also note the underrepresentation of faint LSBGs in the outer parts of the cluster which needs further investigation. A noticeable trend was observed in the $FUV - NUV$ and $NUV - r$ colors as a function of cluster-centric distance, showing that the colors tend to become redder towards the center of the cluster. This trend is consistent with the observations of Venhola et al. (2019b) and Junais et al. (2022). Examining the relation between $NUV - r$ color and the absolute magnitude in the r band, we found that the majority of the NUV-emitting LSBGs are in the blue cloud, with only two LSBGs being red in color, both of which are massive galaxies.

6

Conclusion and Future Perspectives

Currently, the known best description of the universe is the ΛCDM model. This model describes a spatially flat universe with dark energy characterized by a constant energy density, and dark matter consisting of particles with a mass greater than 1 keV. Despite its success in accurately predicting many observations, such as the anisotropies in the cosmic microwave background (CMB) radiation (Planck Collaboration et al. 2016) and the formation of large-scale structures (Davis et al. 1985), the ΛCDM model falls short in addressing some other challenging questions. These include the nature of dark energy and issues on the galaxy scale, collectively referred to as small-scale crises. To test the boundaries of the ΛCDM model and refine our understanding of the universe, we need interesting galaxy systems that can be used to challenge the predictions of our theoretical models.

Strong gravitational lenses and low-surface-brightness galaxies are two such galaxy systems that could be used to test the ΛCDM model and its discrepancies with observations. However, strong gravitational lensing is a rare phenomenon and less than 300 SGLs have been confirmed till now. Similarly, LSBGs are underrepresented in the galaxy catalogues, even after the majority of the galaxy population is hypothesised to be LSBGs. Hence, due to their characteristic nature, these galaxy systems are hard to correctly identify from the enormous amount of data generated in the current and upcoming surveys.

In this thesis, as a first step towards using SGLs and LSBGs to refine our understanding of our universe, I have developed deep learning (DL) based algorithms known as transformers for the identification of SGLs and LSBGs from large-scale surveys. Transformers are DL models introduced in natural language processing that operate on the principle of attention. This mechanism allows the model to selectively focus on relevant parts of the input, similar to how humans can selectively listen to relevant information in a conversation. I analysed the performance of transformers for the analysis of astronomical data for the first time and I have shown that this is a robust algorithm that could be adapted to any domain of astronomy. A detailed discussion of the results and the planned future perspectives are described in Sections 6.1 and 6.2.

6.1 Conclusions

SGL search with transformers

The upcoming large-scale surveys such as LSST and Euclid are expected to identify around 10^5 SGLs by analysing around 10^9 sources (Collett & Auger 2014; Ivezić et al. 2019; Verma et al. 2019; Euclid Collaboration et al. 2022b). In this scenario, relying on semi-automated methods or traditional approaches such as crowd science will not be a practical solution.

Hence, in Thuruthipilly et al. (2022), I explored the possibilities of a new deep-learning technique known as the transformers for the identification of SGLs from large-scale surveys. The simulated dataset created by Metcalf et al. (2019) for the Bologna lens challenge was used to test the performance of transformer models with different architectures and compare their performance with the winning models (mostly CNNs) that participated in the Bologna lens challenge. In addition, it is also shown that the transformers are capable of performing on par with the state-of-the-art CNNs and the best models I have created were able to outperform all the CNN models that participated in the Bologna lens challenge.

Since the transformers have been used for the classification of strong gravitation lenses for the first time, I explored in detail how the hyperparameters of the models enhance their performance. The performance of the models was found to increase with the number of self-attention layers and saturates after a certain number depending on the training dataset. The transformer models is able to filter the relevant features of the image, which can prevent overfitting to an extent and show a more stable performance compared to CNNs. Generally, the transformers are found to be more confident binary classifiers (predicting a classification score of $p \sim 0$ or $p \sim 1$) in comparison to CNNs. In addition, the transformer models trained on a simulated dataset based on KiDS DR 4 are capable of identifying SGL systems from the real data of KiDS DR4 and perform similarly to the CNNs presented in Petrillo et al. (2019). However, it should be noted that the simulated dataset used to train the model does not exactly capture the complexity of the real data and hence directly testing our model without fine-tuning will result in a considerable number of false positives as seen in my follow-up studies; Thuruthipilly et al. (2024a) and Grespan et al. (2024) in which I am the second author.

LSBGs in DES DR 1

It is estimated that the majority of the galaxy population ($\sim$ 50%) resides in the LSB regime (Martin et al. 2019). Nevertheless, they are poorly represented in the current galaxy catalogues. To alleviate this issue, I study the use of transformers for the identification of LSBGs from dark energy survey data release 1 (DES DR 1) and increase the sample size of the LSBGs. For this, I used the 23 790 LSBGs from the DES DR1 identified by Tanoglidis et al. (2021b) to train two different transformer ensemble models capable of separating LSBGs from artefacts in the DES DR 1 (Thuruthipilly et al. 2024b). The trained ensemble models identified 4 083 new LSBGs from the DES DR 1 which was previously missed by Tanoglidis et al. (2021b) and increased the sample size of LSBGs by 17%. Combining the sample of LSBGs from Tanoglidis et al. (2021b) with the newly identified 4 083 LSBGs, the total number of LSBGS in DES DR 1 is increased to 27 873 effectively increasing the number density of LSBGs in DES to 5.57 deg^{-2} from 4.75 deg^{-2}. Hence, extrapolating based on this number density of LSBGs, we are expected to find more than 100 000 and 80 000 LSBGs from LSST and Euclid, respectively. Since the LSST and Euclid are deeper than DES DR 1 by 2 orders of magnitude, the projected estimate of the LSBGs to be found from LSST and Euclid could only be treated as a lower limit.

The distribution of the morphological parameters, such as the Sérsic index and axis ratio, of the total LSBG sample from the DES DR 1 is similar to that of the dwarf galaxy population. This similarity indicates that most galaxies in our sample are likely dwarf galaxies. I divided the LSBG sample based on their $g - i$ color into blue and red galaxy samples, to find that $\sim$ 70% of the LSBGs are blue galaxies, while the remaining $\sim$ 30% are red galaxies. This result aligns with recent findings by Lazar et al. (2024) for dwarf galaxies, supporting the argument that the majority of LSBGs are indeed dwarf galaxies. Furthermore, I investigated the clustering tendencies of LSBGs and discovered that they cluster

more strongly than high-surface-brightness galaxies (HSBGs), driven predominantly by the red LSBGs rather than the blue LSBGs. By cross-matching the total LSBG sample from DES DR1 with the X-ray-selected galaxy cluster catalogue from the ROSAT All-Sky Survey compiled by Xu et al. (2022), a total of 1310 LSBGs were identified to have associated with clusters. Among them, 317 LSBGS could be classified as ultra-diffuse galaxy (UDG) candidates. The LSBGs associated with the cluster showed a weaker trend in color with cluster-centric distance resulting in red LSBGs being located closer to the cluster center compared to those in the outer regions. However, no such significant trend was observed for UDGs. Furthermore, both LSBGs and UDGs tend to be larger in physical size as they are situated farther from the cluster center. These results are promising for the analysis of the LSBGs from the upcoming large-scale surveys such as LSST and Euclid.

LSBGs & UDGs in Abell 194 cluster

The LSST and Euclid are expected to be two orders of magnitude deeper than DES DR1. Therefore, in my next work, I analyzed how well transformer models trained on shallower data could identify LSBGs and UDGs in deeper data. The transformer ensemble models trained on DES DR1 were tested on the data of the Abell 194 cluster (which is also imaged in DES DR 1), imaged with the Hyper-Suprime-Cam (HSC) and has a comparable depth to 10-year LSST data. To account for differences in imaging from DES and HSC, both image data were converted into surface brightness units, ensuring they had similar average pixel values for corresponding spatial regions. After searching with the transformer models and a visual inspection, 171 LSBGs (87 new) among which 28 are UDGs (16 new) were identified. The transformer ensemble models trained on the shallow data were able to identify 93% of the LSBGs without any fine-tuning and missed only 12 LSBGs on a deeper dataset. These results are promising as we can use the existing sample of LSBGs from the shallower datasets to train and identify LSBGs from the upcoming large-scale surveys such as LSST and Euclid and need not resort to semi-automated methods. In this analysis, I improved the masks of the LSBGs by correctly masking small contaminants that may be present on top of the LSBGs. I found that the masks and the subtraction of the local sky background greatly influence the parameters estimated with Sérsic fit.

The newly estimated number of UDGs in Abell 194 is in agreement with the observations in the literature that the number of UDGs in a cluster is linearly correlated with the cluster mass (van der Burg et al. 2017; Mancera Piña et al. 2018; Karunakaran & Zaritsky 2023). Similar to the findings in our previous work (Thuruthipilly et al. 2024b), the Sérsic index distribution of the UDGs is found to be similar to that of dwarf elliptical galaxies indicating that the UDGs might be the tail end of the dwarf galaxy population. Analysing the cluster-centric properties of the LSBGs and UDGs in the Abell 194 cluster showed that the LSBGs and UDGs near the cluster centre tend to have higher Sérsic index than those in the outer parts of the cluster. Similarly, the brighter LSBGs and UDGs are also found to be near the cluster centre. Applying aperture photometry on the available GALEX data of Abell 194, it was also found that 15 LSBGs in Abell 194 emit radiation in both NUV and FUV wavelengths, and 5 LSBGs emit radiation in the NUV wavelength, indicating recent star formation in these galaxies. There was a noticeable trend in the $FUV - NUV$ and $NUV - r$ colors as a function of cluster-centric distance, showing that the colors tend to become redder towards the center of the cluster which is in agreement with the observations of Venhola et al. (2019b) and Junais et al. (2022). Examining the relation between $NUV - r$ color and the absolute magnitude in the r band, it was found that the majority of the NUV-emitting LSBGs are in the blue cloud with only two LSBGs being red in color which both are massive galaxies. The results from this work once again show that

the transformers can be successfully adapted for the analysis of the data from upcoming large-scale surveys such as LSST.

6.2 Future perspectives

Recently, as a follow-up on my results in Thuruthipilly et al. (2022), Grespan et al. (2024) identified 71 high-confidence SGLs, including 44 entirely new ones, by fine-tuning the best-performing models from Thuruthipilly et al. (2022) with known lens candidates in KiDS DR4. However, it should be noted that the systems identified with deep learning models are only SGL candidates and we need spectroscopic redshifts of the source and the lensing galaxy to confirm their status as SGL. Dedicated follow-up spectroscopic surveys such as the Sloan Lens Advanced Camera for Surveys (SLACS; Bolton et al. 2008), Strong Lensing Legacy Survey (SL2S; Sonnenfeld et al. 2013) and ASTRO 3D Galaxy Evolution with Lenses (AGEL; Tran et al. 2022) have been employed in the past to confirm and compile the current known sample of SGLs. The upcoming surveys using the 4-metre Multi-Object Spectroscopic Telescope (4MOST) would increase the confirmed number of SGLs and the spectroscopic data could be used to estimate the stellar velocity dispersion of the lensing galaxy. As noted by Biesiada et al. (2010) and Cao et al. (2015), knowing the redshifts of the lensing galaxy, source galaxy and the velocity dispersion of the lensing galaxy could be used to constrain the cosmological models.

However, obtaining the spectroscopic data would be challenging for the large number of SGLs to be discovered in LSST and Euclid. Hence, as a future follow-up, I am planning to develop ML-based segmentation models which are capable of separating the data of the SGL system into source galaxy, lens galaxy and background which could be used to estimate the lensing parameters such as the Einstein radius and photometric redshifts of the lensing galaxy and source galaxy. By fitting the light profile of the lensing galaxy to a Sérsic profile, one could obtain the surface brightness (μ_e) and the effective radius (R_e) in angular scale which in turn could be used to estimate the velocity dispersion using the fundamental plane relation ($logR_e = alog\mu_e + blog\sigma_0$). However, it should be noted that the coefficients a and b of the fundamental plane relation should be estimated independent of cosmology. This could be achieved by training an artificial neural network on the publicly available supernova Ia data set to predict the distance by knowing the redshift. Knowing the distance to the lensing galaxy independent of cosmology, one could estimate the physical effective radius and calibrate the fundamental plane relation independent of cosmology.

Coming to the LSBGs and UDGs that I have presented in this work, it should be noted that they are only LSBG or UDG candidates as we do not have the spectroscopic redshift of these galaxies to confirm their status as LSBGs and UDGs. However, it is unlikely that the majority of the reported sample are false positives. It is essential to have the redshift information to estimate the physical properties of the LSBGs and UDGs. However, due to their characteristic faintness, the spectroscopic data of LSBGs is hard to obtain. For instance, currently, the known number of UDGs with spectroscopic data is around 50 (Kadowaki et al. 2021). Hence we need dedicated follow-up observations of LSBGs to understand them better. In the future, I am planning to conduct a dedicated large-scale spectroscopic observing campaign to create a large sample of LSBGs with spectroscopic information using the current and upcoming instruments. Examples of these instruments include the multi-unit spectroscopic explorer (MUSE) at the very large telescope (VLT), the multi-object double spectrographs (MODS) at the large binocular telescope (LBT), the keck cosmic web imager (KCWI) at the keck observatory, the multi-espectrógrafo en GTC de alta resolución para astronomía (MEGARA) at the gran telescopio Canaris (GTC), and Bluemuse at the

VLT. In addition, currently, I am a co-investigator of two observing proposals submitted for field blue UDG observations. The first proposal is to observe 20 field-blue UDG candidates from the catalogue of Zaritsky et al. (2022) using the multi-purpose instrument for astronomy at low resolution (MISTRAL) in the Haute-Provence Observatory to estimate the spectroscopic redshifts of the UDGs. The second proposal is a follow-up of 20 field blue LSBGs that I found in Thuruthipilly et al. (2024b) using LBT/MODS to estimate their redshifts and further study their physical properties such as metallicity and star formation surface density of some detected regions. For the sample in both of these observations, we have complimentary data in g, r, z bands UV data from GALEX for some of them. If the proposal gets accepted, we will have the spectroscopic data for $\sim$ 40 UDG candidates which could be used to improve our understanding of the field UDG population.

Most of the LSBGs that I have found from DES Thuruthipilly et al. (2024b) including the sample from Tanoglidis et al. (2021b) are blue. Hence, some fraction of these LSBGs would likely have HI counterparts which could be detected with current and upcoming radio telescopes such as very large array (VLA), the MeerKAT radio and the square kilometre array (SKA). Having well-resolved HI observations will help us to constrain the dynamics of the LSBGs and test different dark matter models similar to the work by Montes et al. (2024). I am also planning to cross-match the sample of LSBGs with the existing catalogues of multi-wavelength observations to enhance our understanding of the LSBG population (e.g., IR: to study dust in LSBGs, sub-mm: to probe molecular gas content in LSBGs, UV: to estimate the star formation histories of LSBGs, X-ray: investigate AGN presence in LSBGs).

One of the interesting observations we found that needs detailed follow-up is that the red LSBGs tend to cluster strongly in comparison to red HSBGs. However, since we do not have the spectroscopic redshifts of the LSBGs, we can not confirm if it is due to the projection effect or any bias in the LSBG sample that we have created. A detailed follow-up study is planned using the marked correlation function instead of the angular correlation function. The idea is to study how different galaxy properties of the LSBGs and HSBGs affect the clustering strength of the LSBGs and HSBGs (Sureshkumar et al. 2021). In addition, we are also planning to obtain robust photometric estimates for LSBGs and to use the photometric redshifts of the LSBGs to investigate if the LSBGs show similar trends in spatial clustering.

Recently, Lazar et al. (2024) observed that the morphological classification of the dwarf galaxies into early-type galaxies (ETGs) and late-type galaxies (LTGs) cannot be correlated to its morphological parameters as seen for the massive galaxies. Similarly, Vega-Ferrero et al. (2021) also found that ML morphological classification models trained on the brighter galaxies can not perform well on fainter galaxies. Since the majority of the LSBG sample, will belong to the dwarf galaxy population, we are planning to split our LSBG sample from DES DR 1 based on their morphology into ETGs (elliptical and lenticular galaxies) and LTGs (spiral and irregular galaxies). Then use the sub-samples with different morphologies to investigate how the morphology of the LSBGs correlates with their morphological properties and how these correlations are different from the HSBGs.

One other major focus in future projects is to investigate the UDG population to test the different UDG formation scenarios. UDGs are an interesting sub-class of LSBGs and various formation mechanisms have been proposed which are still being debated. Currently, UDG formation scenarios predict that the red UDGs will be rare in the field and estimating the red and blue fraction of the observed field UDGs will be able to confirm this prediction. However, it should be noted that the red field dwarf galaxy population is observed to be around 40% (Lazar et al. 2024). If UDGs indeed belong to the subset of dwarf galaxies, exploring the underlying reasons for the distinct color distributions between UDGs and dwarfs would help us understand the UDG formation channels.

Current searches for UDGs have found that the UDGs are found abundant in clusters and they scale in proportion to the halo mass of the cluster (van der Burg et al. 2017; Mancera Piña et al. 2018). Thanks to my collaboration with Prof. Jin Koda, I have dedicated HSC observations of 8 clusters including the Coma and Hydra clusters. I am planning to use the transformer models that I have developed in Thuruthipilly et al. (2024b) to find the LSBG and UDG population in these 8 clusters. The LSBG population found from these clusters can be used to constrain the slope of the luminosity function. In contrast, the UDG population can be used to constrain the relationship between the halo mass of the cluster and the number of UDGs in the cluster. In addition, a collaboration with Prof. Jin Koda is also planned to find globular clusters in the UDGs found from these 8 clusters to estimate the halo mass of the UDGs.

The LSBG and UDG population in the early universe is still unexplored realm due to their lack of detection in large-scale surveys. However, with dedicated observations with powerful telescopes such as the Hubble Space Telescope (HST) and James Web Space Telescope (JWST), we will be able to probe the LSBG population of the early universe and how they have evolved. For instance, Bachmann et al. (2021) used the HST observation to probe the LSBGs in clusters at $z > 1$ and found that the distant UDGs are under-abundant by a factor of ~ 3 compared to local UDGs. Similarly, Carleton et al. (2023) observed the faint population of galaxies in the El Gordo cluster situated at $z \sim 0.87$ with JWST and concluded that multiple formation channels are responsible for UDG formation. However, these observations are done only for small samples of galaxies and we need to extend these analyses to more high-z LSBGs to have statistically confident conclusions. Hence in the future, I am planning to use HST, JWST and extreme large telescope (ELT) to probe the high-z LSBGs and compare their properties with local LSBGs to analyse the evolution of galaxies in the faint regime.

In summary, the SGLs and LSBG population holds the answer to many questions such as the nature of dark matter and dark energy, the contribution to baryons in the local universe by LSBGs and solutions to small-scale problems. The coming age of data will help us to get closer to these answers and the methodologies that I have developed in this thesis will be a stepping stool for this. With the above-mentioned future projects planned I plan to take part in this quest.

Bibliography

Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, ApJS, 182, 543

Abbott, T. M. C., Abdalla, F. B., Allam, S., et al. 2018a, ApJS, 239, 18

Abbott, T. M. C., Abdalla, F. B., Annis, J., et al. 2018b, MNRAS, 480, 3879

Abbott, T. M. C., Adamów, M., Aguena, M., et al. 2021, ApJS, 255, 20

Ackermann, S., Schawinski, K., Zhang, C., Weigel, A. K., & Turp, M. D. 2018, MNRAS, 479, 415

Adami, C., Scheidegger, R., Ulmer, M., et al. 2006, A&A, 459, 679

Aguerri, J. A. L. & González-García, A. C. 2009, A&A, 494, 891

Ahn, C. P., Alexandroff, R., Allende Prieto, C., et al. 2012, ApJS, 203, 21

Ahumada, R., Allende Prieto, C., Almeida, A., et al. 2020, ApJS, 249, 3

Aihara, H., AlSayyad, Y., Ando, M., et al. 2019, PASJ, 71, 114

Aihara, H., AlSayyad, Y., Ando, M., et al. 2022, PASJ, 74, 247

Aihara, H., Armstrong, R., Bickerton, S., et al. 2018, PASJ, 70, S8

Alabi, A. B., Romanowsky, A. J., Forbes, D. A., Brodie, J. P., & Okabe, N. 2020, MNRAS, 496, 3182

Allam, Tarek, J. & McEwen, J. D. 2021, arXiv e-prints, arXiv:2105.06178

Alpher, R. A., Bethe, H., & Gamow, G. 1948, Physical Review, 73, 803

Alzubaidi, L., Zhang, J., Humaidi, A. J., et al. 2021, Journal of Big Data, 8

Amorisco, N. C. & Loeb, A. 2016a, MNRAS, 459, L51

Amorisco, N. C. & Loeb, A. 2016b, MNRAS, 459, L51

Bachmann, A., van der Burg, R. F. J., Fensch, J., Brammer, G., & Muzzin, A. 2021, A&A, 646, L12

Bamford, S. P., Nichol, R. C., Baldry, I. K., et al. 2009, MNRAS, 393, 1324

Barbosa, C. E., Zaritsky, D., Donnerstein, R., et al. 2020, ApJS, 247, 46

Bautista, J. M. G., Koda, J., Yagi, M., Komiyama, Y., & Yamanoi, H. 2023, ApJS, 267, 10

Bay, Y. Y. & Yearick, K. A. 2024, arXiv e-prints, arXiv:2403.01621

Benavides, J. A., Sales, L. V., Abadi, M. G., et al. 2023, MNRAS, 522, 1033

Bernstein, G. M., Abbott, T. M. C., Armstrong, R., et al. 2018, PASP, 130, 054501

Bernstein, G. M., Nichol, R. C., Tyson, J. A., Ulmer, M. P., & Wittman, D. 1995, AJ, 110, 1507

Bertin, E. & Arnouts, S. 1996, A&AS, 117, 393

Bianchi, L., Shiao, B., & Thilker, D. 2017, ApJS, 230, 24

Biesiada, M., Piórkowska, A., & Malec, B. 2010, MNRAS, 406, 1055

Bilicki, M., Jarrett, T. H., Peacock, J. A., Cluver, M. E., & Steward, L. 2014, ApJS, 210, 9

Blandford, R. D. & Narayan, R. 1992, ARA&A, 30, 311

Blanton, M. R., Lupton, R. H., Schlegel, D. J., et al. 2005, ApJ, 631, 208

Boller, T., Freyberg, M. J., Trümper, J., et al. 2016, A&A, 588, A103

Bolton, A. S., Burles, S., Koopmans, L. V. E., et al. 2008, ApJ, 682, 964

Bond, J. R., Efstathiou, G., & Silk, J. 1980, Phys. Rev. Lett., 45, 1980

Bonvin, V., Courbin, F., Suyu, S. H., et al. 2016, MNRAS, 465, 4914

Bonvin, V., Courbin, F., Suyu, S. H., et al. 2017, MNRAS, 465, 4914

Bosch, J., Armstrong, R., Bickerton, S., et al. 2018a, PASJ, 70, S5

Bosch, J., Armstrong, R., Bickerton, S., et al. 2018b, PASJ, 70, S5

Bothun, G., Impey, C., & McGaugh, S. 1997, PASP, 109, 745

Bothun, G. D., Impey, C. D., Malin, D. F., & Mould, J. R. 1987, AJ, 94, 23

Bothun, G. D., Schombert, J. M., Impey, C. D., Sprayberry, D., & McGaugh, S. S. 1993, AJ, 106, 530

Boylan-Kolchin, M., Bullock, J. S., & Kaplinghat, M. 2011, MNRAS, 415, L40

Boylan-Kolchin, M., Springel, V., White, S. D. M., Jenkins, A., & Lemson, G. 2009, MNRAS, 398, 1150

Brout, D., Scolnic, D., Popovic, B., et al. 2022, ApJ, 938, 110

Bull, P., Akrami, Y., Adamek, J., et al. 2016, Physics of the Dark Universe, 12, 56

Bullock, J. S. & Boylan-Kolchin, M. 2017a, ARA&A, 55, 343

Bullock, J. S. & Boylan-Kolchin, M. 2017b, ARA&A, 55, 343

Bullock, J. S., Kolatt, T. S., Sigad, Y., et al. 2001, MNRAS, 321, 559

Burkholder, V., Impey, C., & Sprayberry, D. 2001, AJ, 122, 2318

Buyle, P., De Rijcke, S., Michielsen, D., Baes, M., & Dejonghe, H. 2005, MNRAS, 360, 853

Cañameras, R., Schuldt, S., Suyu, S. H., et al. 2020, A&A, 644, A163

Cabanac, R. A., Alard, C., Dantel-Fort, M., et al. 2007, A&A, 461, 813

Cabrera-Vives, G., Reyes, I., Förster, F., Estévez, P. A., & Maureira, J.-C. 2017, ApJ, 836, 97

Cao, S., Biesiada, M., Gavazzi, R., Piórkowska, A., & Zhu, Z.-H. 2015, ApJ, 806, 185

Caon, N., Capaccioli, M., & D'Onofrio, M. 1993, MNRAS, 265, 1013

Carion, N., Massa, F., Synnaeve, G., et al. 2020, in Computer Vision – ECCV 2020, ed. A. Vedaldi, H. Bischof, T. Brox, & J.-M. Frahm (Cham: Springer International Publishing), 213–229

Carleton, T., Cohen, S. H., Frye, B. L., et al. 2023, ApJ, 953, 83

Carleton, T., Errani, R., Cooper, M., et al. 2019, MNRAS, 485, 382

Chamba, N., Trujillo, I., & Knapen, J. H. 2022, A&A, 667, A87

Chan, T. K., Kereš, D., Wetzel, A., et al. 2018, MNRAS, 478, 906

Chen, P.-C., Tsai, H., Bhojanapalli, S., et al. 2021, in Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (Online and Punta Cana, Dominican Republic: Association for Computational Linguistics), 2974–2988

Chen, Y., Li, R., Shu, Y., & Cao, X. 2019, MNRAS, 488, 3745

Chen, Z., Duan, Y., Wang, W., et al. 2023, in The Eleventh International Conference on Learning Representations

Chianese, M., Coogan, A., Hofma, P., Otten, S., & Weniger, C. 2020, MNRAS, 496, 381

Ciambur, B. C. 2016, Publications of the Astronomical Society of Australia, 33, e062

Clevert, D., Unterthiner, T., & Hochreiter, S. 2016, in 4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings, ed. Y. Bengio & Y. LeCun

Clifton, T., Ferreira, P. G., Padilla, A., & Skordis, C. 2012, Phys. Rep., 513, 1

Cohen, Y., van Dokkum, P., Danieli, S., et al. 2018, ApJ, 96

Collett, T. E. & Auger, M. W. 2014, MNRAS, 443, 969

Conselice, C. J. 2018, Research Notes of the American Astronomical Society, 2, 43

Conselice, C. J., Gallagher, John S., I., & Wyse, R. F. G. 2003a, AJ, 125, 66

Conselice, C. J., O'Neil, K., Gallagher, J. S., & Wyse, R. F. G. 2003b, ApJ, 591, 167

Covone, G., Paolillo, M., Napolitano, N. R., et al. 2009, ApJ, 691, 531

Da Costa, G. S. 1992, in Astronomical Society of the Pacific Conference Series, Vol. 23, Astronomical CCD Observing and Reduction Techniques, ed. S. B. Howell, 90

Dalcanton, J. J., Spergel, D. N., Gunn, J. E., Schmidt, M., & Schneider, D. P. 1997, AJ, 114, 635

Danieli, S., van Dokkum, P., Merritt, A., et al. 2017, ApJ, 837, 136

Davies, A., Serjeant, S., & Bromley, J. M. 2019, MNRAS, 487, 5263

Davis, M., Efstathiou, G., Frenk, C. S., & White, S. D. M. 1985, ApJ, 292, 371

de Amorim, L. B., Cavalcanti, G. D., & Cruz, R. M. 2023, Applied Soft Computing, 133, 109924

de Bernardis, P., Ade, P. A. R., Bock, J. J., et al. 2000, Nature, 404, 955

de Blok, W. J. G. 2010, Advances in Astronomy, 2010, 789293

de Blok, W. J. G., McGaugh, S. S., Bosma, A., & Rubin, V. C. 2001, ApJ, 552, L23

de Jong, J. T. A., Verdoes Kleijn, G. A., Kuijken, K. H., & Valentijn, E. A. 2013, Experimental Astronomy, 35, 25

de Jong, R. S., Bellido-Tirado, O., Brynnel, J. G., et al. 2022, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 12184, Ground-based and Airborne Instrumentation for Astronomy IX, ed. C. J. Evans, J. J. Bryant, & K. Motohara, 1218414

de Vaucouleurs, G. 1948, Annales d'Astrophysique, 11, 247

De Vicente, J., Sánchez, E., & Sevilla-Noarbe, I. 2016, MNRAS, 459, 3078

Dekker, A., Ando, S., Correa, C. A., & Ng, K. C. Y. 2022, Phys. Rev. D, 106, 123026

Deng, J., Dong, W., Socher, R., et al. 2009, in 2009 IEEE Conference on Computer Vision and Pattern Recognition, 248–255

Dey, A., Schlegel, D. J., Lang, D., et al. 2019, AJ, 157, 168

Di Cintio, A., Brook, C. B., Dutton, A. A., et al. 2017, MNRAS, 466, L1

Di Cintio, A., Brook, C. B., Macciò, A. V., Dutton, A. A., & Cardona-Barrero, S. 2019, MNRAS, 486, 2535

Di Valentino, E., Mena, O., Pan, S., et al. 2021, Classical and Quantum Gravity, 38, 153001

Dietterich, T. G. 2000, Multiple Classifier Systems, 1

Disney, M. J. 1976, Nature, 263, 573

Djorgovski, S. & Davis, M. 1987, ApJ, 313, 59

Domingos, P. & Hulten, G. 1999, in Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, 155–164

Domínguez Sánchez, H., Huertas-Company, M., Bernardi, M., et al. 2019, MNRAS, 484, 93

Dosovitskiy, A., Beyer, L., Kolesnikov, A., et al. 2021, in 9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021 (OpenReview.net)

Driver, S. P. 1999, ApJ, 526, L69

Driver, S. P., Liske, J., Cross, N. J. G., De Propris, R., & Allen, P. D. 2005, MNRAS, 360, 81

Drlica-Wagner, A., Bechtol, K., Rykoff, E. S., et al. 2015, ApJ, 813, 109

Du, W., Cheng, C., Zheng, Z., & Wu, H. 2020, The Astronomical Journal, 159, 138

E Greene, J., Greco, J. P., Goulding, A. D., et al. 2022, ApJ, 933, 150

Efron, B. & Gong, G. 1983, The American Statistician, 37, 36

Einasto, J., Saar, E., Kaasik, A., & Chernin, A. D. 1974, Nature, 252, 111

Einstein, A. 1915, Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften, 844

Einstein, A. 1917, Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften, 142

Einstein, A. 1936, Science, 84, 506

Eisenstein, D. J., Weinberg, D. H., Agol, E., et al. 2011, AJ, 142, 72

Erwin, P. 2015, ApJ, 799, 226

Ettori, S. & Balestra, I. 2009, A&A, 496, 343

Euclid Collaboration, Scaramella, R., Amiaux, J., et al. 2022a, A&A, 662, A112

Euclid Collaboration, Scaramella, R., Amiaux, J., et al. 2022b, A&A, 662, A112

Famaey, B. & McGaugh, S. 2013, in Journal of Physics Conference Series, Vol. 437, Journal of Physics Conference Series (IOP), 012001

Fitzpatrick, E. L. 1999, PASP, 111, 63

Flaugher, B., Diehl, H. T., Honscheid, K., et al. 2015, AJ, 150, 150

Forbes, D. A., Dullo, B. T., Gannon, J., et al. 2020, MNRAS, 494, 5293

Freeman, K. C. 1970, ApJ, 160, 811

Friedmann, A. 1924, Zeitschrift fur Physik, 21, 326

Fu, J., Liu, J., Tian, H., et al. 2019, in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (Los Alamitos, CA, USA: IEEE Computer Society), 3141–3149

Furusawa, H., Koike, M., Takata, T., et al. 2018, PASJ, 70, S3

Galaz, G., Frayer, D. T., Blaña, M., et al. 2022, ApJ, 940, L37

Gamow, G. 1946, Physical Review, 70, 572

Gamow, G. 1948, Nature, 162, 680

Gentile, F., Tortora, C., Covone, G., et al. 2021, MNRAS, 510, 500

Girardi, M., Giuricin, G., Mardirossian, F., Mezzetti, M., & Boschin, W. 1998, ApJ, 505, 74

Glorot, X. & Bengio, Y. 2010a, in AISTATS

Glorot, X. & Bengio, Y. 2010b, in JMLR Proceedings, Vol. 9, Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, AISTATS 2010, Chia Laguna Resort, Sardinia, Italy, May 13-15, 2010, ed. Y. W. Teh & D. M. Titterington (JMLR.org), 249–256

Golini, G., Montes, M., Carrasco, E. R., Román, J., & Trujillo, I. 2024, A&A, 684, A99

Graham, A. W. & Driver, S. P. 2005, PASA, 22, 118

Graham, A. W. & Guzmán, R. 2003, AJ, 125, 2936

Greco, J. P., Greene, J. E., Strauss, M. A., et al. 2018, ApJ, 857, 104

Grespan, M., Thuruthipilly, H., Pollo, A., et al. 2024, arXiv e-prints, arXiv:2405.11992

Griffen, B. F., Ji, A. P., Dooley, G. A., et al. 2016, ApJ, 818, 10

Haberzettl, L., Bomans, D. J., & Dettmar, R. J. 2007, A&A, 471, 787

Haghi, H., Kroupa, P., Banik, I., et al. 2019, MNRAS, 487, 2441

Halverson, N. W., Leitch, E. M., Pryke, C., et al. 2002, ApJ, 568, 38

Hanany, S., Ade, P., Balbi, A., et al. 2000, ApJ, 545, L5

Hannon, S., Whitmore, B. C., Lee, J. C., et al. 2023, MNRAS, 526, 2991

Hartley, P., Flamary, R., Jackson, N., Tagore, A. S., & Metcalf, R. B. 2017, MNRAS, 471, 3378

Hawkins, D. M. 2004, Journal of Chemical Information and Computer Sciences, 44, 1, pMID: 14741005

Hayward, C. C., Irwin, J. A., & Bregman, J. N. 2005, ApJ, 635, 827

He, K., Zhang, X., Ren, S., & Sun, J. 2015, in 2015 IEEE International Conference on Computer Vision (ICCV), 1026–1034

He, K., Zhang, X., Ren, S., & Sun, J. 2016, in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778

He, Z., Er, X., Long, Q., et al. 2020, MNRAS, 497, 556

Hewett, P. C. 1982, MNRAS, 201, 867

Hinshaw, G., Larson, D., Komatsu, E., et al. 2013a, ApJS, 208, 19

Hinshaw, G., Larson, D., Komatsu, E., et al. 2013b, ApJS, 208, 19

Hochreiter, S. 1991, Untersuchungen zu dynamischen neuronalen Netzen. Diploma thesis, Institut für Informatik, Lehrstuhl Prof. Brauer, Technische Universität München

Hochreiter, S., Bengio, Y., Frasconi, P., & Schmidhuber, J. 2001, in A Field Guide to Dynamical Recurrent Neural Networks, ed. S. C. Kremer & J. F. Kolen (IEEE Press)

Hornik, K., Stinchcombe, M., & White, H. 1989, Neural Networks, 2, 359

Hosna, A., Merry, E., Gyalmo, J., et al. 2022, Journal of Big Data, 9

Huang, K.-W., Chen, G. C.-F., Chang, P.-W., et al. 2023, in Computer Vision – ECCV 2022 Workshops, ed. L. Karlinsky, T. Michaeli, & K. Nishino (Cham: Springer Nature Switzerland), 143–153

Huang, X., Storfer, C., Gu, A., et al. 2021, ApJ, 909, 27

Huang, X., Storfer, C., Ravi, V., et al. 2020, ApJ, 894, 78

Hubble , E. P. 1925, The Observatory, 48, 139

Hubble, E. 1929, Proceedings of the National Academy of Science, 15, 168

Hubble, E. P. 1926, ApJ, 64, 321

Hwang, S. Y., Sabiu, C. G., Park, I., & Hong, S. E. 2023, J. Cosmology Astropart. Phys., 2023, 075

Impey, C. & Bothun, G. 1997, ARA&A, 35, 267

Impey, C., Bothun, G., & Malin, D. 1988, ApJ, 330, 634

Ivezic, Z., Axelrod, T., Brandt, W. N., et al. 2008, Serbian Astronomical Journal, 176, 1

Ivezić, Ž., Connolly, A. J., VanderPlas, J. T., & Gray, A. 2014, Statistics, Data Mining, and Machine Learning in Astronomy: A Practical Python Guide for the Analysis of Survey Data

Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, ApJ, 873, 111

Jackson, R. A., Martin, G., Kaviraj, S., et al. 2021, MNRAS, 502, 4262

Jacobs, C., Collett, T., Glazebrook, K., et al. 2019, ApJS, 243, 17

Jacobs, C., Glazebrook, K., Collett, T., More, A., & McCarthy, C. 2017, MNRAS, 471, 167

Janiesch, C., Zschech, P., & Heinrich, K. 2021, Electronic Markets, 31, 685

Janowiecki, S., Leisman, L., Józsa, G., et al. 2015, ApJ, 801, 96

Janssens, S. R., Abraham, R., Brodie, J., Forbes, D. A., & Romanowsky, A. J. 2019, ApJ, 887, 92

Jarvis, M. 2015, TreeCorr: Two-point correlation functions, Astrophysics Source Code Library, record ascl:1508.007

Jia, P., Sun, R., Li, N., et al. 2023, AJ, 165, 26

Jones, M. G., Bennet, P., Mutlu-Pakdil, B., et al. 2021, ApJ, 919, 72

Joyce, A., Lombriser, L., & Schmidt, F. 2016, Annual Review of Nuclear and Particle Science, 66, 95

Junais, Boissier, S., Boselli, A., et al. 2022, A&A, 667, A76

Junais, Boissier, S., Epinat, B., et al. 2020, A&A, 637, A21

Junais, Małek, K., Boissier, S., et al. 2023, A&A, 676, A41

Junais, Weilbacher, P. M., Epinat, B., et al. 2024, A&A, 681, A100

Jurić, M., Kantor, J., Lim, K. T., et al. 2017, in Astronomical Society of the Pacific Conference Series, Vol. 512, Astronomical Data Analysis Software and Systems XXV, ed. N. P. F. Lorente, K. Shortridge, & R. Wayth, 279

Kadowaki, J., Zaritsky, D., Donnerstein, R. L., et al. 2021, ApJ, 923, 257

Karunakaran, A. & Zaritsky, D. 2023, MNRAS, 519, 884

Kawanomoto, S., Uraguchi, F., Komiyama, Y., et al. 2018, PASJ, 70, 66

Khan, A., Sohail, A., Zahoora, U., & Qureshi, A. S. 2020, Artif. Intell. Rev., 53, 5455–5516

Kingma, D. P. & Ba, J. 2015, in 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings, ed. Y. Bengio & Y. LeCun

Kniazev, A. Y., Grebel, E. K., Pustilnik, S. A., et al. 2004, AJ, 127, 704

Kobayashi, Y., Nishimichi, T., Takada, M., & Miyatake, H. 2022, Phys. Rev. D, 105, 083517

Koda, J., Yagi, M., Yamanoi, H., & Komiyama, Y. 2015, ApJ, 807, L2

Koekemoer, A. M. 2019, in American Astronomical Society Meeting Abstracts, Vol. 234, American Astronomical Society Meeting Abstracts #234, 222.02

Komiyama, Y., Obuchi, Y., Nakaya, H., et al. 2018, PASJ, 70, S2

Koo, D. C. & Szalay, A. S. 1984, ApJ, 282, 390

Koopmans, L. V. E. 2006, in EAS Publications Series, Vol. 20, EAS Publications Series, ed. G. A. Mamon, F. Combes, C. Deffayet, & B. Fort, 161–166

Koopmans, L. V. E., Treu, T., Bolton, A. S., Burles, S., & Moustakas, L. A. 2006, ApJ, 649, 599

Krizhevsky, A., Sutskever, I., & Hinton, G. E. 2012, in Advances in Neural Information Processing Systems 25, ed. F. Pereira, C. J. C. Burges, L. Bottou, & K. Q. Weinberger (Curran Associates, Inc.), 1097–1105

Kuhn, V., Guo, Y., Martin, A., et al. 2023, arXiv e-prints, arXiv:2312.12389

Kuijken, K., Heymans, C., Dvornik, A., et al. 2019, A&A, 625, A2

La Marca, A., Iodice, E., Cantiello, M., et al. 2022a, A&A, 665, A105

La Marca, A., Iodice, E., Cantiello, M., et al. 2022b, A&A, 665, A105

Landy, S. D. & Szalay, A. S. 1993, ApJ, 412, 64

Lanusse, F., Ma, Q., Li, N., et al. 2017, MNRAS, 473, 3895

Laudato, E. & Salzano, V. 2023, European Physical Journal C, 83, 402

Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, arXiv e-prints, arXiv:1110.3193

Lazar, I., Kaviraj, S., Watkins, A. E., et al. 2024, MNRAS, 529, 499

Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. 1998, Proceedings of the IEEE, 86, 2278

Lee, C. H., Hodges-Kluck, E., & Gallo, E. 2020, MNRAS, 497, 2759

Lee, J.-W. 2021, Journal of Korean Physical Society, 78, 873

Leisman, L., Haynes, M. P., Janowiecki, S., et al. 2017, ApJ, 842, 133

Lemaître, G. 1927, Annales de la Société Scientifique de Bruxelles, 47, 49

Lenzen, F., Schindler, S., & Scherzer, O. 2004, A&A, 416, 391

Li, R., Napolitano, N. R., Tortora, C., et al. 2020, ApJ, 899, 30

Li, T., Collett, T. E., Krawczyk, C. M., & Enzi, W. 2024, MNRAS, 527, 5311

Lim, S., Côté, P., Peng, E. W., et al. 2020, ApJ, 899, 69

Liutkus, A., Cífka, O., Wu, S., et al. 2021, in Proceedings of Machine Learning Research, Vol. 139, Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event, ed. M. Meila & T. Zhang (PMLR), 7067–7079

Ma, C.-P. & Fry, J. N. 2000, ApJ, 543, 503

Magnier, E. A., Schlafly, E., Finkbeiner, D., et al. 2013, ApJS, 205, 20

Magro, D., Zarb Adami, K., DeMarco, A., Riggi, S., & Sciacca, E. 2021, MNRAS, 505, 6155

Mallat, S. 2016, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374, 20150203

Mancera Piña, P. E., Aguerri, J. A. L., Peletier, R. F., et al. 2019, MNRAS, 485, 1036

Mancera Piña, P. E., Golini, G., Trujillo, I., & Montes, M. 2024, arXiv e-prints, arXiv:2404.06537

Mancera Piña, P. E., Peletier, R. F., Aguerri, J. A. L., et al. 2018, MNRAS, 481, 4381

Marleau, F. R., Habas, R., Poulain, M., et al. 2021, A&A, 654, A105

Marshall, P. J., Verma, A., More, A., et al. 2016, MNRAS, 455, 1171

Martin, D. C., Fanson, J., Schiminovich, D., et al. 2005, ApJ, 619, L1

Martin, G., Kaviraj, S., Laigle, C., et al. 2019, MNRAS, 485, 796

Mateo, M. L. 1998, ARA&A, 36, 435

McConnachie, A. W. 2012, AJ, 144, 4

Mcculloch, W. & Pitts, W. 1943, Bulletin of Mathematical Biophysics, 5, 127

McGaugh, S. S. 1996, MNRAS, 280, 337

McGaugh, S. S. & Bothun, G. D. 1994, AJ, 107, 530

McGaugh, S. S., de Blok, W. J. G., Schombert, J. M., Kuzio de Naray, R., & Kim, J. H. 2007, ApJ, 659, 149

McGaugh, S. S., Schombert, J. M., & Bothun, G. D. 1995, AJ, 109, 2019

McKean, J., Jackson, N., Vegetti, S., et al. 2015, in Advancing Astrophysics with the Square Kilometre Array (AASKA14), 84

Merritt, A., van Dokkum, P., Danieli, S., et al. 2016, ApJ, 833, 168

Merritt, D. 2017, Studies in the History and Philosophy of Modern Physics, 57, 41

Metcalf, R. B., Meneghetti, M., Avestruz, C., et al. 2019, A&A, 625, A119

Metcalf, R. B. & Petkova, M. 2014, MNRAS, 445, 1942

Mihos, J. C., Durrell, P. R., Ferrarese, L., et al. 2015, ApJ, 809, L21

Mihos, J. C., Harding, P., Feldmeier, J. J., et al. 2017, ApJ, 834, 16

Minchin, R. F., Disney, M. J., Parker, Q. A., et al. 2004, MNRAS, 355, 1303

Miyazaki, S., Komiyama, Y., Kawanomoto, S., et al. 2018, PASJ, 70, S1

Mo, H., van den Bosch, F. C., & White, S. 2010, Galaxy Formation and Evolution

Mo, H. J., McGaugh, S. S., & Bothun, G. D. 1994, MNRAS, 267, 129

Montes, M. 2019, arXiv e-prints, arXiv:1912.01616

Montes, M., Trujillo, I., Karunakaran, A., et al. 2024, A&A, 681, A15

Moore, B., Ghigna, S., Governato, F., et al. 1999, ApJ, 524, L19

Moore, B., Katz, N., Lake, G., Dressler, A., & Oemler, A. 1996, Nature, 379, 613

Morganson, E., Gruendl, R. A., Menanteau, F., et al. 2018, PASP, 130, 074501

Morrissey, P., Conrow, T., Barlow, T. A., et al. 2007, ApJS, 173, 682

Müller, O., Jerjen, H., & Binggeli, B. 2018, A&A, 615, A105

Navarro, J. F., Eke, V. R., & Frenk, C. S. 1996, MNRAS, 283, L72

Navarro, J. F., Ludlow, A., Springel, V., et al. 2010, MNRAS, 402, 21

Neilsen, E. J., , Annis, J. T., et al. 2019, Dark Energy Survey's Observation Strategy, Tactics, and Exposure Scheduler, Tech. rep.

Neuschaefer, L. W., Windhorst, R. A., & Dressler, A. 1991, ApJ, 382, 32

Niu, Z., Zhong, G., & Yu, H. 2021, Neurocomputing, 452, 48

O'Neil, K. & Bothun, G. 2000, ApJ, 529, 811

O'Neil, K. L. 1997, PhD thesis, University of Oregon

Ostriker, J. P., Peebles, P. J. E., & Yahil, A. 1974, ApJ, 193, L1

Parmar, N., Ramachandran, P., Vaswani, A., et al. 2019, in Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada, 68–80

Peacock, J. A. & Smith, R. E. 2000, MNRAS, 318, 1144

Pearson, J., Li, N., & Dye, S. 2019, MNRAS, 488, 991

Pearson, W. J., Suelves, L. E., Ho, S. C. C., et al. 2022, A&A, 661, A52

Peebles, P. J. E. 1980, The large-scale structure of the universe

Peebles, P. J. E. 1982, ApJ, 263, L1

Peebles, P. J. E. & Hauser, M. G. 1974, ApJS, 28, 19

Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2002, AJ, 124, 266

Penzias, A. A. & Wilson, R. W. 1965, ApJ, 142, 419

Pérez-Carrasco, M., Cabrera-Vives, G., Martinez-Marin, M., et al. 2019, PASP, 131, 108002

Pérez-Montaño, L. E., Rodriguez-Gomez, V., Cervantes Sodi, B., et al. 2022, MNRAS, 514, 5840

Perivolaropoulos, L. & Skara, F. 2022, New A Rev., 95, 101659

Perlmutter, S., Aldering, G., Goldhaber, G., et al. 1999, ApJ, 517, 565

Persic, M. & Salucci, P. 1992, MNRAS, 258, 14P

Petrillo, C. E., Tortora, C., Chatterjee, S., et al. 2019, MNRAS, 482, 807

Petrillo, C. E., Tortora, C., Chatterjee, S., et al. 2017, MNRAS, 472, 1129

Petrillo, C. E., Tortora, C., Vernardos, G., et al. 2019, MNRAS, 484, 3879

Pimentel, Ó., Estévez, P. A., & Förster, F. 2023, AJ, 165, 18

Pinochet, J. & Van Sint Jan, M. 2018, Physics Education, 53, 055003

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, A&A, 594, A13

Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2020, A&A, 641, A6

Popolo, A. D. & Le Delliou, M. 2017, 43pp 19 figs , invited review, accepted by Galaxies

Poulain, M., Marleau, F. R., Habas, R., et al. 2021, MNRAS, 506, 5494

Prole, D. J., van der Burg, R. F. J., Hilker, M., & Davies, J. I. 2019, MNRAS, 488, 2143

Prole, D. J., van der Burg, R. F. J., Hilker, M., & Spitler, L. R. 2021, MNRAS, 500, 2049

Rhodes, J., Nichol, R. C., Aubourg, É., et al. 2017, ApJS, 233, 21

Riess, A. G., Casertano, S., Yuan, W., Macri, L. M., & Scolnic, D. 2019, ApJ, 876, 85

Riess, A. G., Filippenko, A. V., Challis, P., et al. 1998, AJ, 116, 1009

Rines, K., Geller, M. J., Kurtz, M. J., & Diaferio, A. 2003, AJ, 126, 2152

Roberts, M. S. & Rots, A. H. 1973, A&A, 26, 483

Robertson, H. P. 1935, ApJ, 82, 284

Robitaille, T. & Bressert, E. 2012, APLpy: Astronomical Plotting Library in Python, Astrophysics Source Code Library, record ascl:1208.017

Rojas, K., Savary, E., Clément, B., et al. 2022, A&A, 668, A73

Román, J. & Trujillo, I. 2017a, MNRAS, 468, 4039

Román, J. & Trujillo, I. 2017b, MNRAS, 468, 4039

Román, J., Trujillo, I., & Montes, M. 2020, A&A, 644, A42

Rosa, R. 2020, Anais da Academia Brasileira de Ciencias, 93, e20200861

Rubin, V. C., Ford, W. K., J., & Thonnard, N. 1978, ApJ, 225, L107

Rubin, V. C., Ford, W. K., J., & Thonnard, N. 1980, ApJ, 238, 471

Russakovsky, O., Deng, J., Su, H., et al. 2015, Int. J. Comput. Vis., 115, 211

Saburova, A. S., Chilingarian, I. V., Kasparova, A. V., et al. 2021, MNRAS, 503, 830

Saburova, A. S., Chilingarian, I. V., Kulier, A., et al. 2023, MNRAS, 520, L85

Sales, L. V., Navarro, J. F., Peñafiel, L., et al. 2020, MNRAS, 494, 1848

Sammut, C. & Webb, G. I. 2010, Encyclopedia of Machine Learning

Sandage, A. & Binggeli, B. 1984, AJ, 89, 919

Sato, H. & Takahara, F. 1980, Progress of Theoretical Physics, 64, 2029

Schaefer, C., Geiger, M., Kuntzer, T., & Kneib, J.-P. 2018, A&A, 611, A2

Schaye, J., Crain, R. A., Bower, R. G., et al. 2015, MNRAS, 446, 521

Schlafly, E. F. & Finkbeiner, D. P. 2011, ApJ, 737, 103

Schlafly, E. F., Finkbeiner, D. P., Jurić, M., et al. 2012, ApJ, 756, 158

Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525

Schmidgall, S., Achterberg, J., Miconi, T., et al. 2023, ArXiv, abs/2305.11252

Sérsic, J. L. 1963, Boletin de la Asociacion Argentina de Astronomia La Plata Argentina, 6, 41

Sevilla-Noarbe, I., Bechtol, K., Carrasco Kind, M., et al. 2021, ApJS, 254, 24

Shankaranarayanan, S. & Johnson, J. P. 2022, General Relativity and Gravitation, 54, 44

Shull, J. M., Smith, B. D., & Danforth, C. W. 2012, ApJ, 759, 23

Simon, J. D. 2019, ARA&A, 57, 375

Simon, J. D. & Geha, M. 2007, ApJ, 670, 313

Simonyan, K. & Zisserman, A. 2015, in 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings, ed. Y. Bengio & Y. LeCun

Singh, P. R., Zaritsky, D., Donnerstein, R., & Spekkens, K. 2019, AJ, 157, 212

Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163

Somalwar, J. J., Greene, J. E., Greco, J. P., et al. 2020, ApJ, 902, 45

Sonnenfeld, A., Treu, T., Gavazzi, R., et al. 2013, ApJ, 777, 98

Spergel, D. N., Bean, R., Doré, O., et al. 2007, ApJS, 170, 377

Sprayberry, D., Impey, C. D., Bothun, G. D., & Irwin, M. J. 1995, AJ, 109, 558

Springel, V., Wang, J., Vogelsberger, M., et al. 2008, MNRAS, 391, 1685

Strateva, I., Ivezić, Ž., Knapp, G. R., et al. 2001, AJ, 122, 1861

Su, H., Yi, Z., Liang, Z., et al. 2024, MNRAS, 528, 873

Su, J., Ahmed, M., Lu, Y., et al. 2024, Neurocomputing, 568, 127063

Sureshkumar, U., Durkalec, A., Pollo, A., et al. 2021, A&A, 653, A35

Tan, A., Nguyen, D. T., Dax, M., Nießner, M., & Brox, T. 2021, in Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021, 9799–9807

Tanoglidis, D., Ćiprijanović, A., & Drlica-Wagner, A. 2021a, Astronomy and Computing, 35, 100469

Tanoglidis, D., Drlica-Wagner, A., Wei, K., et al. 2021b, ApJS, 252, 18

Tempel, E., Kipper, R., Tamm, A., et al. 2016, A&A, 588, A14

Thuruthipilly, H., Grespan, M., & Zadrożny, A. 2024a, in American Institute of Physics Conference Series, Vol. 3061, American Institute of Physics Conference Series, 040003

Thuruthipilly, H., Junais, Pollo, A., et al. 2024b, A&A, 682, A4

Thuruthipilly, H., Zadrozny, A., Pollo, A., & Biesiada, M. 2022, A&A, 664, A4

Tollerud, E. J., Boylan-Kolchin, M., & Bullock, J. S. 2014, MNRAS, 440, 3511

Tonry, J. L., Stubbs, C. W., Lykke, K. R., et al. 2012, ApJ, 750, 99

Tran, K.-V. H., Harshan, A., Glazebrook, K., et al. 2022, AJ, 164, 148

Tremaine, S. & Gunn, J. E. 1979, Phys. Rev. Lett., 42, 407

Treu, T. 2010, ARA&A, 48, 87

Treu, T. & Marshall, P. J. 2016, A&A Rev., 24, 11

Tsujikawa, S. 2013, Classical and Quantum Gravity, 30, 214003

Tulin, S. & Yu, H.-B. 2018, Phys. Rep., 730, 1

Tully, R. B. & Fisher, J. R. 1977, A&A, 54, 661

Tuominen, T., Nevalainen, J., Tempel, E., et al. 2021, A&A, 646, A156

van der Burg, R. F. J., Hoekstra, H., Muzzin, A., et al. 2017, A&A, 607, A79

van der Burg, R. F. J., Muzzin, A., & Hoekstra, H. 2016, A&A, 590, A20

van Dokkum, P., Abraham, R., Brodie, J., et al. 2016, ApJ, 828, L6

van Dokkum, P., Danieli, S., Cohen, Y., et al. 2018, Nature, 555, 629

van Dokkum, P. G., Abraham, R., Merritt, A., et al. 2015a, ApJ, 798, L45

van Dokkum, P. G., Romanowsky, A. J., Abraham, R., et al. 2015b, ApJ, 804, L26

Vaswani, A., Shazeer, N., Parmar, N., et al. 2017, in Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA, 5998–6008

Vega-Ferrero, J., Domínguez Sánchez, H., Bernardi, M., et al. 2021, MNRAS, 506, 1927

Venhola, A., Peletier, R., Laurikainen, E., et al. 2019a, A&A, 625, A143

Venhola, A., Peletier, R., Laurikainen, E., et al. 2019b, A&A, 625, A143

Venhola, A., Peletier, R. F., Salo, H., et al. 2022, A&A, 662, A43

Verma, A., Collett, T., Smith, G. P., Strong Lensing Science Collaboration, & the DESC Strong Lensing Science Working Group. 2019, arXiv e-prints, arXiv:1902.05141

Viel, M., Becker, G. D., Bolton, J. S., & Haehnelt, M. G. 2013, Phys. Rev. D, 88, 043502

Vilalta, R. 2018, Journal of Physics: Conference Series, 1085, 052014

Vogelsberger, M., Genel, S., Springel, V., et al. 2014, MNRAS, 444, 1518

Voges, W., Aschenbach, B., Boller, T., et al. 1999, A&A, 349, 389

Walsh, D., Carswell, R. F., & Weymann, R. J. 1979, Nature, 279, 381

Wang, X., Kondratyuk, D., Christiansen, E., et al. 2022, in International Conference on Learning Representations

Wang, X., Liu, Y., & Xin, H. 2021, Structures, 32, 2279

Wei, W., Huerta, E. A., Whitmore, B. C., et al. 2020, MNRAS, 493, 3178

White, S. D. M., Davis, M., & Frenk, C. S. 1984, MNRAS, 209, 27P

Wittmann, C., Lisker, T., Ambachew Tilahun, L., et al. 2017, MNRAS, 470, 1512

Wong, K. C., Suyu, S. H., Chen, G. C. F., et al. 2020, MNRAS, 498, 1420

Wortsman, M., Ilharco, G., Gadre, S. Y., et al. 2022, in ICML, 23965–23998

Xing, Y., Yi, Z., Liang, Z., et al. 2023, ApJS, 269, 59

Xu, J.-L., Zhu, M., Yu, N., et al. 2023, ApJ, 944, L40

Xu, W., Ramos-Ceja, M. E., Pacaud, F., Reiprich, T. H., & Erben, T. 2022, A&A, 658, A59

Yagi, M., Kashikawa, N., Sekiguchi, M., et al. 2002, AJ, 123, 66

Yagi, M., Koda, J., Komiyama, Y., & Yamanoi, H. 2016, ApJS, 225, 11

Yang, X. 2020, Journal of Physics: Conference Series, 1693, 012173

Yasuda, N., Fukugita, M., Narayanan, V. K., et al. 2001, AJ, 122, 1104

Yi, Z., Li, J., Du, W., et al. 2022, MNRAS, 513, 3972

Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. 2014, in Advances in Neural Information Processing Systems, ed. Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, & K. Weinberger, Vol. 27 (Curran Associates, Inc.)

Yu, J., Wang, Z., Vasudevan, V., et al. 2022, Trans. Mach. Learn. Res., 2022

Zaritsky, D., Donnerstein, R., Karunakaran, A., et al. 2022, ApJS, 261, 11

Zehavi, I., Zheng, Z., Weinberg, D. H., et al. 2005, ApJ, 630, 1

Zhang, H., Goodfellow, I., Metaxas, D., & Odena, A. 2019, in Proceedings of Machine Learning Research, Vol. 97, Proceedings of the 36th International Conference on Machine Learning, ed. K. Chaudhuri & R. Salakhutdinov (PMLR), 7354–7363

Zhao, H., Jia, J., & Koltun, V. 2020, in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 10073–10082

Zheng, Z., Berlind, A. A., Weinberg, D. H., et al. 2005, ApJ, 633, 791

Zhong, G. H., Liang, Y. C., Liu, F. S., et al. 2008, MNRAS, 391, 986

Zibetti, S. 2008, in Dark Galaxies and Lost Baryons, ed. J. I. Davies & M. J. Disney, Vol. 244, 176–185

Zwicky, F. 1933, Helvetica Physica Acta, 6, 110

Zwicky, F. 1937, Physical Review, 51, 290