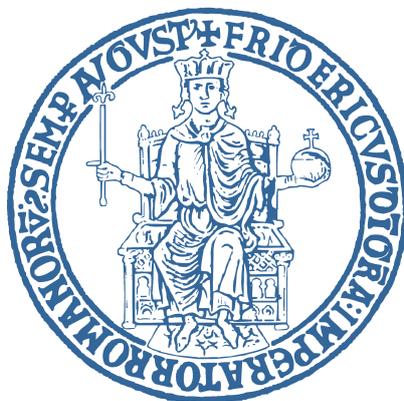


Università degli Studi di Napoli
“Federico II”



Scuola Politecnica e delle Scienze di Base
Area Didattica di Scienze Matematiche Fisiche e Naturali
Dipartimento di Fisica “Ettore Pancini”

Laurea Magistrale in Fisica

Finding gravitational lenses using transfer learning

Relatori:

Prof. Giuseppe Longo
Dott. Crescenzo Tortora
Dott. Stefano Cavuoti

Candidato:

Gianluca Sasanelli
Matricola:
N94000785

A.A. 2022/2023

Two things fill the mind with ever-increasing wonder and awe, the more often and the more intensely the mind of thought is drawn to them: the starry heavens above me and the moral law within me”
- Immanuel Kant

Contents

1	Introduction	1
2	Gravitational Lensing	3
2.1	Lensing theory	4
2.1.1	Thin Lens Approximation	5
2.1.2	Lens Equation	6
2.2	Image Magnification	7
2.3	Lens Models	8
2.3.1	Point Mass Lens	9
2.3.2	Extended Mass Distribution	9
2.3.3	Singular Isothermal Sphere	10
2.3.4	Singular Isothermal Ellipsoid	11
3	Neural Networks and ULISSE	13
3.1	Neural Networks	13
3.1.1	Training	15
3.2	Feature Extractor	16
3.3	k-Nearest Neighbors Algorithm	18
3.4	ULISSE	22
4	Data	23
4.1	The Kilo Degree Survey (KiDS)	23
4.2	The Dataset	25
4.2.1	Luminous red galaxy sample	25
4.2.2	Simulation	26
5	Results	31
5.1	Tuning of the parameters	31
5.2	Statistics	34
5.2.1	Statistics of simulation parameters	36
5.3	Quasi-realistic case	38
6	Discussion and Conclusions	41
A	Relativistic Deflection Angle	43
B	Sérsic profile	45
C	Point Spread Function	47

List of Figures

2.1	The Molten Ring	3
2.2	Image of the lensed quasar Q0957+561	4
2.3	Deflection of a point particle with mass m by the massive deflector M in the Newtonian framework	5
2.4	Typical geometry of gravitational lensing	6
2.5	Schematic representation of magnification of lensed source	8
2.6	Image of the lensed quasar UZC J224030.2+03213	12
3.1	The Sigmoid function	14
3.2	Diagram of an MLP	15
3.3	Convolutional layer	17
3.4	Effects of different convolutional kernels	18
3.5	Representation of a graph	19
3.6	Visual representation of a tree data structure	20
3.7	Representation of a K-D tree	21
3.8	Schematic representation of the dataset partition using the ball-tree algorithm	21
3.9	Overview of ULISSE	22
4.1	Illustration of the spatial distribution of survey tiles unveiled in KiDS-ESO-DR4	24
4.2	Three examples of the bright galaxy sample	25
4.3	Three examples of LRG sample	26
4.4	Three-band magnitudes (g, r, i) of the Sérsic source.	28
4.5	RGB simulated images of gravitationally lensed objects	29
4.6	Final distributions of the simulation parameters	29
5.1	Examples of the Dataset	32
5.2	Representation of a lensed galaxy in RGB and individual bands	32
5.3	Average accuracy of positive queries	32
5.4	Average accuracy of negative queries	33
5.5	Accuracy of ULISSE in the n th most frequent image within the similar of positive queries	35
5.6	First fifty images for frequency within the list of similar of the positive queries	35
5.7	Relationship between various simulation parameters and the accuracy of ULISSE. The plot shows the average accuracy over 30 data points.	37
5.8	Ten examples of real lens candidates found by Petrillo et al. 2019b	38
5.9	Score of real lens candidates vs. accuracy	38
5.10	Example of how ULISSE works	39
5.11	Accuracy vs. number of candidates	40
B.1	Sérsic surface brightness profiles and Sérsic aperture magnitude profiles	46

C.1	convolution of the surface brightness of the astronomical object with the PSF of the telescope	47
C.2	Normalized intensity profiles of Gaussian and Moffat PSF functions	48
C.3	PSF anisotropy in KiDS	49

List of Tables

4.1	KiDS observing strategy	24
4.2	Parameters range and distribution for the simulated mass distribution of the lens (SIE)	27
4.3	Parameters for the Sérsic source	27
4.4	PSF parameters	28
5.1	Score counts	40

LIST OF TABLES

Chapter 1

Introduction

The amount of data collected from digital surveys has grown exponentially in the last twenty years and will continue to grow in the next decade, reaching the scales of the exabyte with the Square Kilometer Array (Dewdney et al. 2009) and the Vera C. Rubin Observatory (Ivezić et al. 2019). The extensive datasets produced by these types of facilities introduce complexities when it comes to data analysis and processing, prompting the creation of a novel field in Astrophysics known as Astrominformatics (Brescia et al. 2017). This field uses data mining and machine learning (ML) methodologies for data analysis in the astronomical context (Longo et al. 2019). These tools can quickly make accurate inference or, in general, discover complex relationships from large quantities of data. Machine learning techniques can be roughly divided in two paradigms: Supervised ML, which relies on extensive datasets of labelled data to train algorithms and Unsupervised ML, which clusters data on the base of density and vicinity in the feature space, and thus does not require a dataset of labels. This makes the Unsupervised paradigm completely data driven and could provide ways to expand our understanding of the field (e.g, Boroson and Green 1992; Protopapas et al. 2006; Vanderplas and Connolly 2009; Baron and Poznanski 2017; Castro-Ginard et al. 2018; Razim et al. 2021). On the other hand, supervised methods are easier to understand and to deal with. In fact, most of the works in astrophysics are made with supervised methods (e.g, Weir et al. 1995; Connolly et al. 1995; Collister and Lahav 2004; Brescia et al. 2013; Djorgovski et al. 2016; Delli Veneri et al. 2019; Mahabal et al. 2019; Wenzl et al. 2021; Gentile et al. 2023).

For instance, the best results in image recognition (He et al. 2015) are obtained with Convolutional Neural Networks (CNNs), a supervised method (LeCun et al. 1989b;a; 1998; LeCun et al. 2015). After the performance of AlexNet (Krizhevsky et al. 2012), CNNs have been successfully applied to many scientific domains (e.g, Jarrett et al. 2009; Cai et al. 2020), including Astronomy. However, CNNs require extensive labelled datasets (at least 10^5 images) to be trained effectively. In the astronomical field, this can not only be costly but sometimes an impossible goal. Consider the application of CNN to strong gravitational lenses recognition. Large databases of gravitational lenses reach ~ 700 elements (e.g, the MasterLens project¹), which is at least two orders of magnitude less than what is required for the effective training of such models. Moreover, even if one collected the right amount of confirmed gravitational lenses, additional complexities arise, such as that each lens type may not contain adequate sampling. Consequently, a CNN might develop a bias, improving its ability to recognize more common types of lenses while underperforming in the identification of the less probable ones. For these reasons, simulations serve as a crucial tool for the training of CNNs for this task. This is obviously a suboptimal way to train CNN because simulations are based on our current understanding of the Universe and therefore may contain systemic inaccuracies. Hence, an exploration of alternative methods becomes interesting. A valuable machine learning technique, particularly in scenarios of limited data availability, is Transfer Learning (Pan and Yang 2009). This technique involves

¹<https://test.masterlens.org/>

deploying a pre-trained algorithm that has been initially developed for a distinct task (for a review of its application in Astrophysics, refer to [Vilalta et al. 2019](#)). The rationale behind this approach arises from the observation that humans exhibit enhanced efficiency in learning new tasks by using knowledge acquired from related tasks ([Jankowski et al. 2013](#)). ULISSE (aUto-matic Lightweight Intelligent System for Sky Exploration) algorithm ([Doorenbos et al. 2022](#)), a transfer learn method, has been developed for object retrieval within astronomical datasets. It is a one-shot method capable of retrieving objects closely related to a query, it makes use of a pre-trained CNN in order to decompose the images in features and then uses a similarity-search algorithm to obtain the nearest neighbors in the feature-space previously created. This method acts as a complement to traditional CNNs, offering a viable alternative when extensive labelled are not available. While CNNs can obtain the highest accuracy with sufficient training, ULISSE exchanges accuracy for broader applicability. Additionally, ULISSE can be especially useful for the identification of rare, thus interesting, astrophysical objects, for which CNNs may not have received (adequate) training.

This thesis studies ULISSE’s performance in the specific task of identifying strong gravitational lensed galaxies. Strong Gravitational lensing (SGL) consists in the deflection of light rays propagating from a distant source by the gravitational field of a massive object (e.g, galaxy or cluster), creating multiple images, arcs, and rings around the deflector (also named “lens”). The scientific potential of this phenomenon is important for three main reasons: Firstly, strong lensing can magnify distant sources, hence allowing the observation of objects that otherwise would have been too small to be resolved ([Impellizzeri et al. 2008](#); [Swinbank et al. 2009](#); [Richard et al. 2010](#); [Deane et al. 2013](#); [Treu et al. 2015](#); [Kelly et al. 2018](#); [Welch et al. 2022](#)). Secondly, the observables depend only on the projected two-dimensional mass distribution of the lens and are independent of the luminosity or composition of the lens. Therefore, it is an ideal tool to study the internal composition of galaxies and especially dark matter properties (e.g, [Treu and Koopmans 2004](#); [Koopmans 2005](#); [Tortora et al. 2010](#); [Auger et al. 2010](#); [Vegetti et al. 2014](#); [Shajib et al. 2021](#)). Lastly, if the background source is time-dependent (e.g., supernovas or quasar), the delay between the arrival time of photons of multiple images can be used to obtain a measurement of the Hubble constant H_0 (see more in [Birrner et al. 2022](#)), independent of the local distance ladder ([Refsdal 1964](#); [Wong et al. 2020](#)). For a more detailed review of strong lensing applications, refer to [Treu 2010](#) and [Blandford and Narayan 1992](#). Currently, the scientific community is investing considerable effort in developing robust algorithms (usually CNNs) tailored for the recognition of strongly lensed galaxies in preparation for new data obtained by telescopes, like ESA’s *Euclid* satellite ([Laureijs et al. 2011](#)), the *Chinese space station* ([Gong et al. 2019](#)) and the *Vera Rubin Observatory*. These facilities will retrieve $\sim 10^5$ strongly lensed galaxies ([Collett 2015](#)) within a population of billions galaxies. Historically, identified lenses have been predominantly massive elliptical galaxies with redshifts falling within the range of $0.1 < z < 1$ ([Shajib et al. 2022](#)). However, with the introduction of these new telescopes, a significant paradigm shift is expected, not only with the substantial increase in raw data but also with the potential for more high-redshift images.

The thesis is structured as follows. In Chapter 2, the lensing theory is briefly summarised. In Chapter 3, a theoretical overview of Neural Networks, Convolutional Neural Networks, k-nearest neighbors search, and ULISSE is provided. The focus will be on examining the fundamental components that constitute ULISSE. In Chapter 4, the details of the datasets and the simulation procedure followed to obtain lensed objects are outlined. In Chapter 5, the results obtained by testing the algorithm on simulated and real datasets are summarised. In Chapter 6, the results are discussed, and possible applications of the algorithm when applied to other datasets are forecasted. Finally, our conclusions and future perspectives of the algorithm are summarised.

Chapter 2

Gravitational Lensing



Figure 2.1: The Molten Ring. *Hubble Space Telescope image* © ESA/Hubble & NASA, S. Jha, 2020.

The concept of light deflection by a gravitational field was first suspected by Isaac Newton in the 18th century ([Newton 1704](#)). Precise calculations were first published by Soldner ([Soldner 1801](#)), however, all calculations made with the framework of Newtonian physics will rely on the assumption that light particles have non-zero mass, which is in contrast with our current understanding of light. On the contrary, General Relativity predicts that mass-less particles, as well as mass particles, are affected by gravity. In fact, Gravitational Lensing is one of the great consequences of Einstein's General Theory of Relativity ([Einstein 1916](#)), and it was the first test of its validity ([Dyson et al. 1920](#)). Yet, at the time, it was unclear whether gravitational

lensing would have a future as a field in astrophysics on its own. Einstein (Einstein 1936) himself rejected this concept, primarily due to his belief that the deflection caused by stars would produce multiple images with an angular separation so small that they would be impossible to resolve.

However, the first who understood the potential of the gravitational lensing was the Swiss astronomer Fritz Zwicky in 1937 (Zwicky 1937a;b). He realized that also galaxies can act both as lenses and as sources, which means that the probability of the interaction between two of them is higher than in the case of stars.

The real proof of Zwicky's idea arrived 40 years later. The detection of the multiple imaged quasar (Q0957+561, Walsh et al. 1979) is the first detection of gravitational lensing on cosmic scale. The images have roughly the same spectra, meaning that the source is the same.

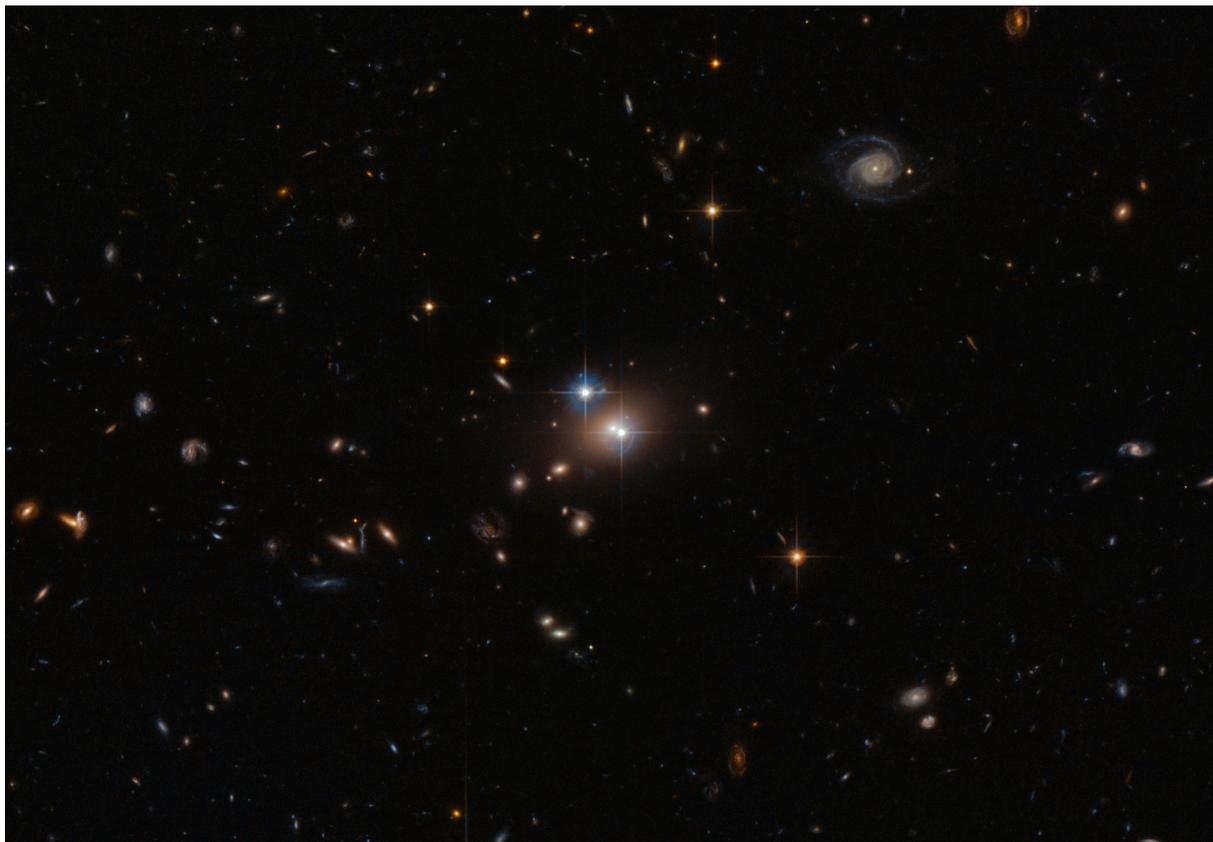


Figure 2.2: HST image of Q0957+561. The two blue objects in the center are lensed images of a quasar at redshift $z = 1.41$ and are separated by 5.7 arcsecond. The two objects have the same redshift and spectra.

Nowadays, gravitational lensing is a field on its own. There are more than 700 confirmed lensed images¹, and the phenomenon is now regarded by the scientific community as one of the most powerful tools to study the composition and geometry of the Universe.

In this chapter, we briefly review the theory of the strong gravitational lensing. We proceed to outline the fundamental principles of this phenomenon, its primary observational consequences and different parametric models for the deflector.

2.1 Lensing theory

The first estimate of the angle of deflection can be obtained using Classical Physics. Assuming the following statements:

¹A gravitational lens confirmation is obtained only through spectroscopic measurement

- Light is made of point particles of mass m moving at the speed of light c ,
- The deflector is a point particle of mass $M \gg m$,
- Small angle approximation²;

and with some algebra, it is possible to compute the deflection angle :

$$\hat{\alpha}_{classic} = \frac{2GM}{c^2 b}. \quad (2.1)$$

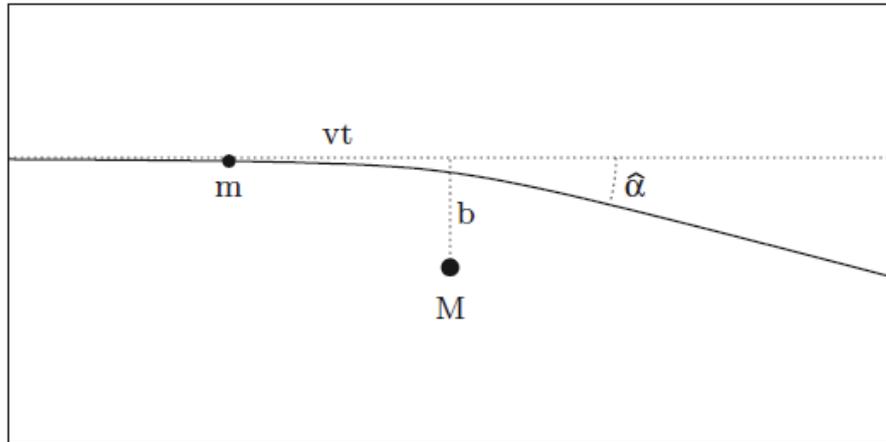


Figure 2.3: Representation of the deflection of a point particle with mass m by the massive deflector M in the Newtonian framework. Original source: [Keeton 2014](#)

The geometry of the mechanics is described in Fig.2.3, where b is the impact parameter and G is the gravitational constant. Note that the deflection angle is independent of the mass of the “light particle”, therefore it applies to arbitrarily low masses. Computing the same angle in the framework of General Relativity gives a deflection angle that is double (see Appendix A). Usually, it is preferable to work in the Newtonian framework and add the factor 2 by hand, obtaining:

$$\hat{\alpha} = \frac{4GM}{c^2 b}. \quad (2.2)$$

2.1.1 Thin Lens Approximation

The previous equation 2.2 can be generalized to any mass distribution with the “thin lens approximation”. Gravitational lensing depends on the total distribution of mass along the Line of Sight (LoS) from the observer to the source. However, the size of the deflector is much smaller than the cosmological distances between the deflector and the observer and the deflector and the source. Therefore, in the thin lens approximation, it is sufficient to consider the projection of the three-dimensional mass density of the lens onto the 2-dimensional plane, called *lens plane*. Similarly, the three-dimensional light distribution of the source can be reduced to the projection onto the *source plane*, namely the surface brightness.

The lens is defined by his mass distribution given by $\rho(\theta', z)$, where θ' defines the position in the lens plane and z represents the position along the LoS. The projection results in a surface density of:

$$\Sigma(\theta') = \int \rho(\theta', z) dz. \quad (2.3)$$

²also known as weak Gravitational Field approximation)

With these considerations, strong lensing can be re-framed as a transformation between the two-dimensional observed coordinated of a certain light ray θ (the image plane) to the two-dimensional coordinated which would be observed in the absence of the deflector β (the source plane).

2.1.2 Lens Equation

The transformation between these two planes is characterized by the famous lens equation. Its derivation begins by defining D_l and D_s as the angular diameter distances from the observer to the lens and to the source, respectively, and D_{ls} as the distance from the lens to the source.³

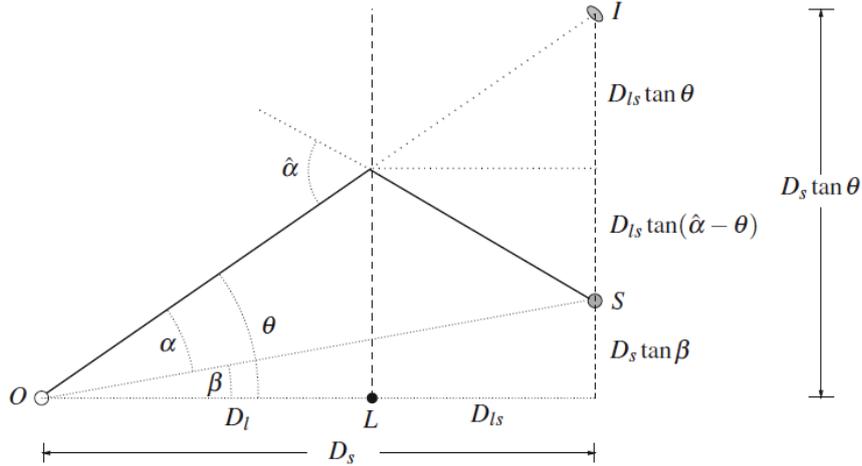


Figure 2.4: Typical geometry of gravitational lensing. The light rays emitted by the source S , observed at an angle β , are deflected by the lens L and produce the lensed image I seen by the observer O at an angle θ . The angular difference between S and I is given by $\hat{\alpha}$, which is the deflection angle. Original source: Keeton 2014.

Assuming a Euclidean geometry (at least in the vicinity of the lens), we can write the following equation:

$$D_s \tan \beta = D_s \tan \theta - D_{ls} \tan \hat{\alpha}(\theta). \quad (2.4)$$

Using the small-angle approximation⁴ and defining the reduced deflection angle as:

$$\alpha(\theta) = \frac{D_{ls}}{D_s} \hat{\alpha}(\theta), \quad (2.5)$$

the **lens equation** read as:

$$\beta = \theta - \alpha(\theta). \quad (2.6)$$

The equation represents the transformation from the image plane (θ) to the source plane (β). In other words, given a certain observed image at θ and knowing the reduced deflection angle $\alpha(\theta)$, we can compute the angular position of the source. There are three things to note here:

- The transformation is achromatic. It transforms only the shape of the source.
- The transformation conserves surface brightness.
- The equation is non-linear, thus a single source can be associated with multiple images.

³Note that, because the geometry of the Universe is not Euclidean, these angular distances do not add up in a simple way. Hence, $D_{ls} \neq D_s - D_l$

⁴ $\tan \theta \approx \sin \theta \approx \theta$

Following usual conventions in physics, we can define the *lens potential* as a 2-D potential, so that its gradient will be the reduced deflection angle:

$$\alpha = \nabla\psi. \quad (2.7)$$

2.2 Image Magnification

As seen before, SGL affects the shape of the source while conserving the surface brightness. This property is named as magnification effect. Mathematically, the magnification of a point source, with coordinates θ , is defined by the magnification matrix:

$$|\mu| = \left| \frac{d\Omega_I}{d\Omega_S} \right| = \left| \frac{\theta d\theta}{\beta d\beta} \right|, \quad (2.8)$$

where $d\Omega_I$ and $d\Omega_S$ are elements of solid angle in the image and source planes. Note that the Jacobian of the lens equation is proportional to the inverse of the magnification. This tensor determines how the shape of a source is transformed by the deflection, and we refer to it as amplification tensor:

$$\mathcal{A} = \frac{\partial\beta}{\partial\theta} = \begin{Bmatrix} 1 - k - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - k + \gamma_1 \end{Bmatrix}, \quad (2.9)$$

where:

$$k = \frac{1}{2} \left(\frac{\partial^2\psi}{\partial\theta_1^2} + \frac{\partial^2\psi}{\partial\theta_2^2} \right), \quad (2.10)$$

$$\gamma_1 = \frac{1}{2} \left(\frac{\partial^2\psi}{\partial\theta_1^2} - \frac{\partial^2\psi}{\partial\theta_2^2} \right), \quad (2.11)$$

$$\gamma_2 = \frac{\partial^2\psi}{\partial\theta_1\partial\theta_2}. \quad (2.12)$$

Each one of these components depends on the second order derivative of the lens potential. k is related to focusing light, thus it is called convergence. γ_1 and γ_2 are related to the distortion of the image, hence they are known as *shear*.

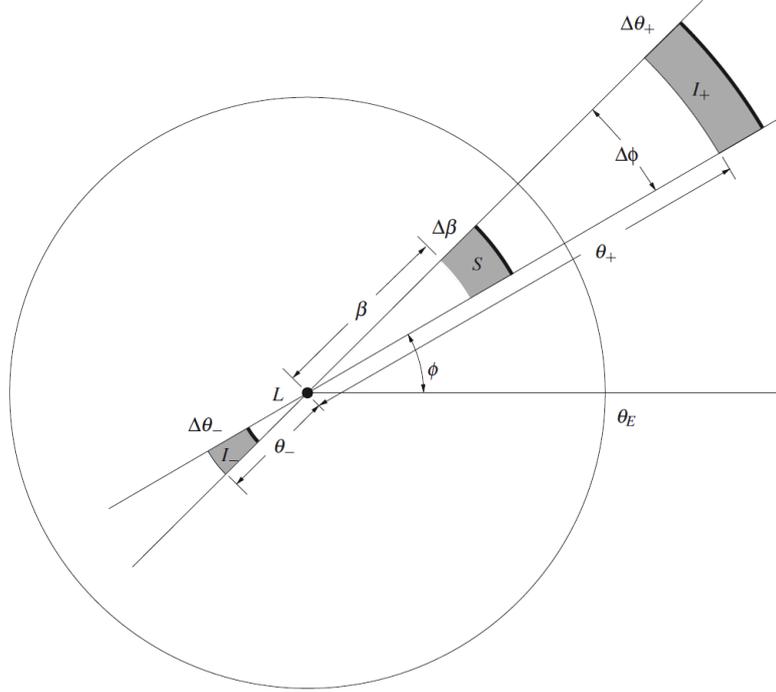


Figure 2.5: Schematic representation of magnification of lensed sources. The source S at angular distance β from the lens L produces images I_{\pm} at distances θ_{\pm} in the plane of the image. The infinitesimal area of S ($d\Omega_S$) is represented as $\beta\Delta\beta\Delta\phi$, while that of I_{\pm} ($d\Omega_I$) is $\theta_{\pm}\Delta\theta_{\pm}\Delta\phi$. When $\beta = 0$ an “Einstein ring” is formed. Original source: [Congdon and Keeton 2018](#)

These components describe what happens for sources that are small enough to be constant across the lens. For extended sources, the magnification depends both on the surface brightness and on the amplification tensor. Values of θ that cause magnification to diverge are referred to as critical curves, while their counterparts in the source plane are known as caustics.

Using Eq. (2.9), one can write the magnification as function of convergence and shear as:

$$\mu(\theta) = \frac{1}{\det(\mathcal{A})} \frac{1}{[(1 - k(\theta))^2 - \gamma_1^2(\theta) - \gamma_2^2(\theta)]}. \quad (2.13)$$

We name the case in which $\mu(\theta) < 0$ as de-magnification, while the case $\mu(\theta) > 0$ is named magnification. The magnification ratio between two different positions θ_1 and θ_2 in the source plane, named *flux ratio*, is an important observable of the phenomenon.

Magnification is commonly employed to study high- z sources as said in Chapter 1. Furthermore, the convergence enables us to employ a quantitative approach for distinguishing between two distinct frames of Gravitational Lensing:

- When $k \ll 1$, there is a negligible distortion of the source, indicating that we are in the weak lensing regime.
- Conversely, when $k \gg 1$, distortion becomes dominant, defining the strong lensing regime.

2.3 Lens Models

When examining galaxy lenses, it is necessary to account for the distributed mass, typically through a parameterized model. Here, we review three fundamental models:

1. Point Mass Lens

2. Singular Isothermal Sphere
3. Singular Isothermal Ellipsoid

2.3.1 Point Mass Lens

The simplest model of a deflector is a point mass. Combining Eq. (2.2) and Eq. (2.5), the scaled deflection angle can be written as:

$$\alpha = \frac{4GM}{c^2} \frac{D_{ls}}{D_l D_s} \frac{1}{\theta}. \quad (2.14)$$

Here we used the substitution $b = D_l \theta$, which is valid assuming the small angle approximation (see Fig. 2.4). Defining the Einstein radius as:

$$\theta_E = \left(\frac{4GM}{c^2} \frac{D_{ls}}{D_l D_s} \right)^{1/2}, \quad (2.15)$$

we can solve the lens equation for θ using β as a parameter, the results is:

$$\theta_{\pm} = \frac{1}{2} [\beta \pm (\beta^2 + 4\theta_E^2)^{1/2}]. \quad (2.16)$$

For $\beta = 0$ the solutions are $\theta_{\pm} = \pm\theta_E$. In this scenario the observers, lens and source lie on the same line, therefore there is a perfect symmetry and the image takes the form of a circle and θ_E provides the angular size of it, this case is named *Einstein ring*. It is worth noting that $\beta = 0$ is a caustic point and the critical curves associated represent a solution that is an Einstein ring. In the general case $\beta \neq 0$, we have two solutions. One inside the Einstein ring and one located outside. Moreover, with some algebra, we can obtain the following formula:

$$M = \frac{c^2}{4G} \frac{D_l D_s}{D_{ls}} |\theta_+ \theta_-|. \quad (2.17)$$

In this formula is enclosed the potential of gravitational lensing to study the Universe. From observing two lensed images, and knowing the distances involved, we can obtain the mass of the lens.

2.3.2 Extended Mass Distribution

For stars, the point mass approximation is a reasonable model. However, modelling galaxies requires additional parameters beyond mass. Due to the thin-lens approximation, the important feature of an extended mass distribution (i.e., galaxies or clusters) is the projected surface mass density, Σ , defined in Eq. (2.3). In this case, the deflection angle read as:

$$\alpha(\theta) = \frac{1}{\pi} \int \frac{\Sigma(\theta')}{\Sigma_{crit}} \frac{\theta - \theta'}{(\theta - \theta')^2} d\theta', \quad (2.18)$$

where:

$$\Sigma(\theta') = \int \rho(z, \theta') dz, \quad (2.19)$$

$$\Sigma_{crit} = \frac{c^2}{4\pi G} \frac{D_l D_s}{D_{ls}}. \quad (2.20)$$

We refer to Σ_{crit} as the critical surface density. It is worthy considering that taking the divergence of α and using Eq. (2.7), we obtain:

$$\nabla \cdot \alpha = \nabla^2 \psi = 2 \frac{\Sigma}{\Sigma_{crit}}. \quad (2.21)$$

Comparing with Eq. (2.10) we note that:

$$k = \frac{\Sigma}{\Sigma_{crit}} = \frac{1}{2} \nabla^2 \psi. \quad (2.22)$$

The Poisson equation connects the lens potential with the convergence, which is practically the dimensionless surface mass density.

2.3.3 Singular Isothermal Sphere

The simplest parametric model of a galaxy is the Singular Isothermal Sphere (SIS). The galaxy is treated as a spherically-symmetrical object in hydrostatic equilibrium. The pressure emerging from the collision of the particles is balanced with the gravitational force. Assuming that locally the galaxy follows the ideal gas law, the pressure is then a function of velocity dispersion, σ , and mass density, ρ :

$$P(r) = \sigma^2 \rho(r). \quad (2.23)$$

In order to remove P from the equation, we make use of the hydrostatic equilibrium equation, which reads as:

$$\nabla P = -\rho \nabla \Psi, \quad (2.24)$$

where Ψ is the gravitational potential of the galaxy. Combining both equations yields the density profile:

$$\rho(r) = \frac{\sigma^2}{2\pi G} \frac{1}{r^2}. \quad (2.25)$$

Despite the many assumptions, this geometry is commonly employed in astrophysics because it predicts flat circular velocity in the outskirts of the galaxies, which corresponds well with empirical findings. It is important to highlight that the density profile diverges for $r \rightarrow 0$ (this explains the term *Singular*). To address this problem, it is usually employed a slightly modified model called Nonsingular Isothermal Sphere (NIS, see more in [Schneider et al. 1992](#)).

To study the SIS as a lens model, we must first compute the deflection angle. Combining Eq. (2.25) and Eq. (2.18) yields the deflection angle:

$$\alpha = 4\pi \left(\frac{\sigma}{c}\right)^2 \frac{D_{ls}}{D_s}. \quad (2.26)$$

Note that it is independent of the position of the light ray. In this situation, the Einstein radius is equal to the deflection angle $\theta_E = \alpha$. The lens equation will depend on the direction of the deflection, hence two possible scenarios exist. The first one is for $\theta > 0$, which yields:

$$\theta = \beta + \theta_E. \quad (2.27)$$

The second one is when $\theta < 0$:

$$\theta = \beta - \theta_E. \quad (2.28)$$

Now, as before, the solution will depend on the value of β . Here, we can distinguish three cases:

- $\beta = 0$ (i.e., when the source is aligned with the deflector) the image forms a ring with angular radius θ_E called Einstein Ring,
- $0 < \beta \leq \theta_E$: two images,
- $\beta > \theta_E$: one image.

The most robust quantity we can infer using gravitational lensing is the mass enclosed in the Einstein radius. With some algebra, we can write $M(\theta_E)$ as a function of the velocity dispersion or circular velocity of the lens:

$$M(\theta_E) = \frac{\sigma^2 \pi D_l \theta_E}{G}. \quad (2.29)$$

It becomes clear that the Einstein radius can be used as a natural angular scale in the lensing theory. The typical angular separation of images is of the order of $2\theta_E$. Therefore, the development of telescopes for observing these images must take this value as reference of angular resolution, which should be lower. Moreover, the Einstein radius expresses the lensing strength of a deflector and its well determined by strong lens models. As a consequence, the mass enclosed within the Einstein radius can be measured with low error (1-2 %). Using Eq. (2.3.1) and typical values for the masses and distances, we can compute common values for the Einstein radius:

$$\begin{cases} \theta_E \approx (0.9 \text{ mas}) \left(\frac{M}{M_\odot} \right)^{\frac{1}{2}} \left(\frac{D}{10 \text{ kpc}} \right)^{-\frac{1}{2}} \\ \theta_E \approx (0.9 \text{ arcsec}) \left(\frac{M}{10^{11} M_\odot} \right)^{\frac{1}{2}} \left(\frac{D}{10 \text{ Gpc}} \right)^{-\frac{1}{2}} \end{cases} \quad (2.30)$$

The first expression is more useful when considering a stellar-scale lens within the Galaxy, known as microlensing. Conversely, the second one considers an extra-galactic galaxy-scale or cluster-scale lensing event.

2.3.4 Singular Isothermal Ellipsoid

The SIS model does not predict most of the observed images of the strong lensing phenomenon, like the so-called “quads” (i.e., four images of the source). This is due to the fact that massive galaxies are elliptical (and not spherical). This highlights the need to consider models that take into account an elliptical geometry, like the Spherical Isothermal Ellipsoid (SIE, [Kormann et al. 1994](#)).

The core hypothesis of the SIE model is that the deflector is described by elliptical isodensity curves with an axis-ratio $0 < q < 1$ and a position angle γ . Defining the ellipse coordinate as $\xi = (\theta_1^2 + \theta_2^2/q^2)^{1/2}$, the Eq. (2.25) now transform as:

$$\rho(\xi) = \frac{\sigma^2}{2\pi G} \frac{1}{\xi}. \quad (2.31)$$

With the SIE it can be difficult to compute the integral 2.18. However, the lens equation can be computed analytically. The most important solution of this type of model is the quads, an example is shown in Fig. (2.6). It occurs when the source and deflector align perfectly along the LoS.

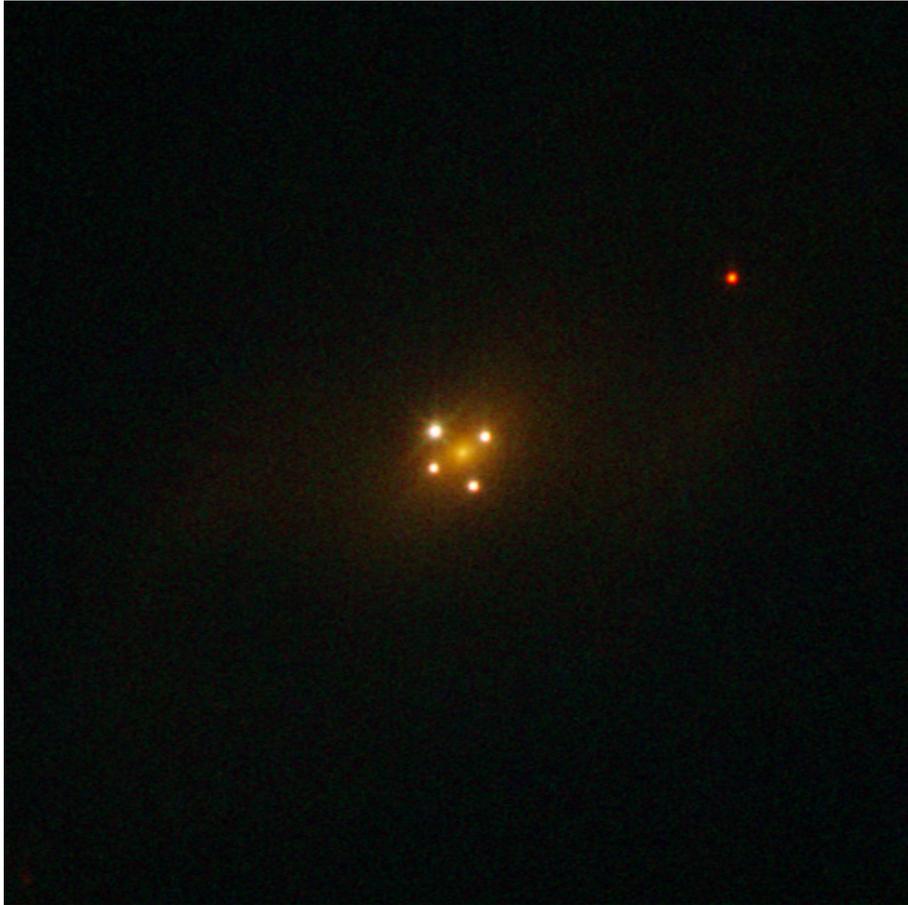


Figure 2.6: This NASA/ESA Hubble Space Telescope picture of UZC J224030.2+03213 depicts a strong lensed quasar. Positioned at the center of the image is the deflector, whereas the other four light sources are from the same quasar at redshift of approx. $z = 0.8$. The alignment between the two objects is within 0.05 arcseconds. This particular observation stands as one of the finest instances of quads, underscoring the capabilities of the Hubble Space Telescope. Credit for this image is attributed to (ESA/Hubble & NASA)

The SIE model has three free parameters. Two of these are connected to the geometry of the lens: the position angle and the axis-ratio. The third parameter is related to the mass content, which is described by the Einstein radius. A more complex model can be defined by adding an external shear component to the SIE. This considers the impact of neighboring galaxies in the lens environment or adjusts for the distribution of matter along the LoS (Keeton et al. 1997). In this case, the external shear introduces two more independent variables to the model: the intensity of shear and its orientation.

Chapter 3

Neural Networks and ULISSE

In the machine learning field, Artificial Neural Networks (ANNs) represent a class of models inspired by the complex networks found in the human brain, particularly the nervous system. These models are often presented as a complex system of interconnected neurons exchanging information. One specific variant of ANN is the Convolutional Neural Network (CNN), renowned for its ability to effectively break down an image into an array of features, which then are used for classification of the image.

ULISSE uses a pre-trained CNN to decompose astronomical images into feature vectors, hence removing the last layers which are usually employed for classification. Subsequently, a similarity search algorithm, such as k-Nearest Neighbors (kNN), is applied to identify the closest vector to a given query. It's important to note that this approach aligns with the principle of Transfer Learning, as the CNN was not originally trained for the classification of astronomical images.

The primary aim of this chapter is to provide the essential theoretical foundation for comprehending ULISSE. As such, the chapter briefly explores neural network theory (Sections 3.1 and 3.2), kNN theory (Section 3.3) and ULISSE's architecture (Section 3.4)

3.1 Neural Networks

ANNs aim to approximate a specific function $f(x)$. In our study, this function takes as input the pixels of an image and returns the classification label indicating whether the image depicts a lensed galaxy (1) or not (0). The ability of the human mind to perform this task suggests the existence of such a function, even though it is not clear how to represent it using simple addition and multiplication. Exists many ways to approximate a certain function $f(x)$; one consists in starting with a set of functions $f(x; \theta)$ parameterized by a vector of *model parameters* θ_ν , which can then be tuned in order to find the best approximation of the initial function (this phase is referred to as *training*). ANN, are a particular set of functions that mimic the functioning of the human brain. It is composed by a high number of computational nodes (neurons) linked to each other by the so-called “connections” (akin to synapses in biology). The connections are responsible for transmitting information from one neuron to another, and we will refer to them as “weights”, whereas the artificial neuron performs two operations:

- Linear sum: It first performs the linear sum of incoming signals x_j weighed by W_{ij} and biased by b_i in order to obtain the preactivation value ψ_i :

$$\psi_i = b_i + \sum_{j=1}^{n_{in}} W_{ij} x_j \quad \text{for } i = 1, \dots, n_{out} ; \quad (3.1)$$

- Activation: The neuron *fires* or not according to the outcome of a nonlinear scalar function

σ called “activation function” which take as input ψ_i :

$$\phi_i = \sigma(\psi_i). \quad (3.2)$$

The activation function is independent of the model parameters, hence of the training process. For instance, a typical activation function is the *Sigmoid* σ :

$$\sigma(x) = \frac{1}{1 + e^{-\psi}} = \frac{1}{2} + \frac{1}{2} \tanh \frac{x}{2}. \quad (3.3)$$

The function is nearly linear over a small range of inputs value around 0, hence preserving the information of the preactivation value. Outside the range, it fast converges to saturation: with output of 0 for $x \rightarrow -\infty$ and 1 for $x \rightarrow \infty$. The function is plotted in Fig. 3.1.

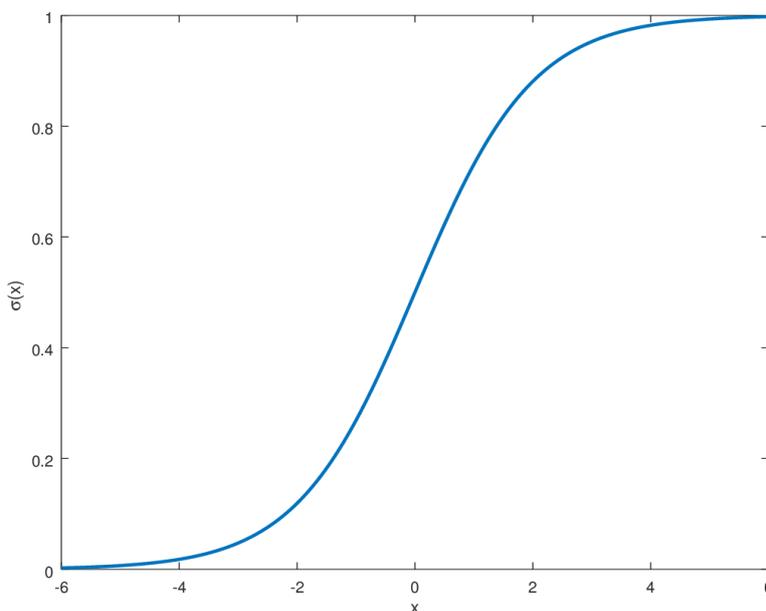


Figure 3.1: The Sigmoid function

In a typical ANN, neurons are organized into layers. The n_{in} neurons of each layer serve as inputs, weighted by the associated connections and biased by b_i , to the n_{out} neurons in the next layer. Therefore, a layer is a function parameterized by the biases b_i , a weight matrix W_{ij} , and an activation function $\sigma(\psi)$.

Many layers can be stacked serially to augment the model parameters, hence approximating better the desired function $f(x)$. This principle is called “multi-layer perceptron” (MLP, Rosenblatt 1963; Rumelhart et al. 1986) and it is at the base of the famous *deep learning* (Hinton et al. 2006) paradigm¹.

The layers of an MLP can be divided in three general categories:

1. Input layer: The layer that receives the external data
2. Hidden layers: These are the intermediate layers which construct the approximation of the classification function
3. Output layer: Final layer that produce the ultimate result of the neural networks. Typically, it is a classification label

¹Note, that the activation function of the first MLP is a Heaviside step function(θ) instead of modern activation function

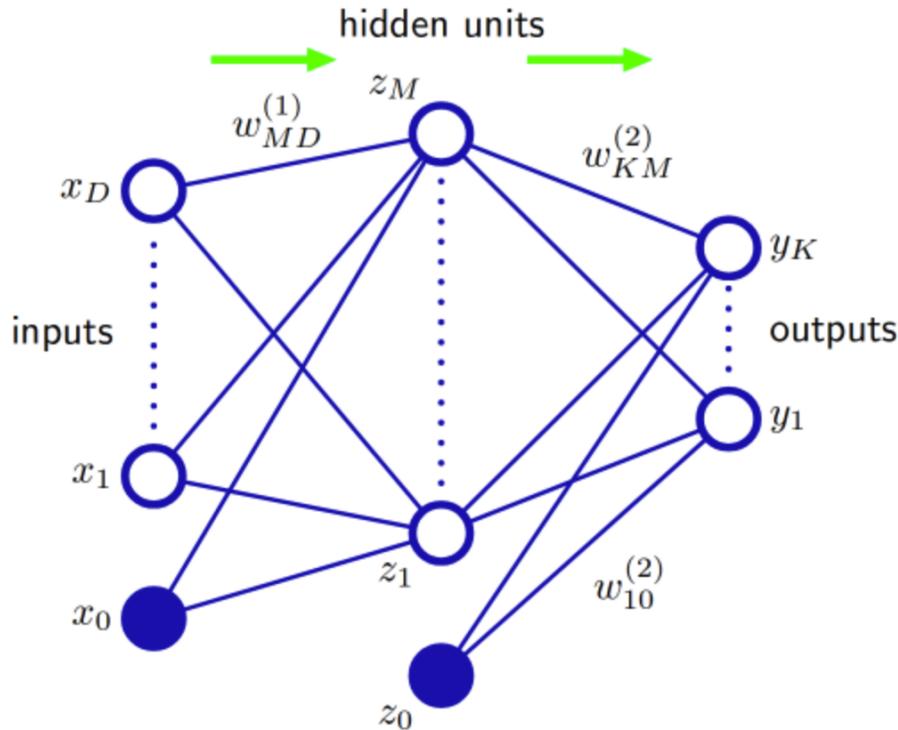


Figure 3.2: Diagram of an MLP. The input, hidden and output variables are represented by nodes, the weights parameters by the links connecting the nodes and the bias parameters are the links coming from the input and hidden variables x_0 and z_0 . Original source : [Bishop 2006](#)

Per the Universal Approximation Theorem ([Hornik et al. 1989](#)), it is established that an opportune Multi-Layer Perceptron can approximate any given function. This result implies that any limitations in achieving success in practical applications can be attributed to inadequate learning, an insufficient number of hidden units, or the absence of a deterministic relationship between the input and target variables.

3.1.1 Training

Training is the process of updating the model parameters. The size of a neural network, and thus the number of its parameters, is enormous, reaching the order of 10^{11} for ChatGPT-3 ([Brown et al. 2020](#)). This implies that the tuning of the parameters must be computationally fast.

The idea of the training, in the Supervised framework, is to use a large dataset of labelled input (e.g., in CNN typically datasets of at least 10^5 images are needed to perform a proper training); then evaluating the output of the ANN against the “ground truth” and modifying the parameters based on the “dissimilarity”. More precisely, the mathematical representation of dissimilarity is the so-called loss function, or also known as the objective function:

$$\mathcal{L}(\phi(x; \theta), y). \quad (3.4)$$

It must be a globally continuous and differentiable function that represents the difference between model-predicted labels and ground truth². The rationale for the introduction of the loss function

²A common loss function is the mean-squared error (MSE):

$$MSE = \frac{1}{N} \sum_{i=1}^N y - y' \quad (3.5)$$

is that the training can now be re-framed as an optimization problem, with the objective of finding the global minimum of the loss function, which corresponds to the best set of weights for the specific task at hand. There are many approaches to optimization problems in mathematics. However, the most frequent in the field of neural network optimization is gradient-descent (and its variations). At each iteration of training, we will update the weights following the equation:

$$\theta_{\nu+1} = \theta_{\nu} - \gamma \nabla \mathcal{L}(\theta_{\nu}). \quad (3.6)$$

The logic of the equation is that the negative gradient of the loss function will have the direction of the local minimum of the function in the parameters' space. To find the local minimum, we can update the weights with the negative of the gradient multiplied by a hyperparameter called “learning rate”, denoted with γ in the previous equation.

However, the goal is to find the *global* minimum, and this is why today exists many versions of this framework with many “updating” equations that include other terms to make the optimization, not only better but also more efficient.

3.2 Feature Extractor

By specifying a loss function, employing an optimization procedure, and using a comprehensive dataset with labeled information, artificial neural networks can independently and effectively “learn” different tasks. Nevertheless, when it comes to processing images, ANNs fall short (ref). The reason is that images contain vast amounts of data, which, if left unfiltered, would exponentially increase the quantity of the ANN’s parameters. This is precisely why Convolutional Neural Networks were developed.

Convolutional neural networks are a special class of ANN specialized in image classification. The idea behind CNNs is to compress the image in a feature vector and then use a simple classifier on this vector. Therefore, for our purposes, it is useful to divide the CNN into two distinct blocks:

- Feature extractor
- Classifier

The feature extractor is the novelty of the CNN, it transforms the input image through convolutional layers, activation functions, and pooling layers, in an array of “features”. Usually, the first layer of neurons which processes the image is the convolutional layer. It applies small small-sized convolutional kernel on the image to detect local features over the whole image and produces the so-called feature maps. The weights are “shared” by the input image (i.e., they remain constant in the convolution). Thus, it reduces by many orders of magnitude the number of parameters needed by the model (see Fig. 3.3 and Fig. 3.4). After it, Activation functions are used. They, as usual, add nonlinearity to the model³. At last, some pooling layer may be necessary. It combines regions of the feature map through statistical operation (e.g., the maximum of the region) effectively reducing the size of the feature map and hence of the model parameters.

³Typically the *Relu* (rectified linear unit) is used in CNNs:

$$f(x) = \frac{x + |x|}{2} \quad (3.7)$$

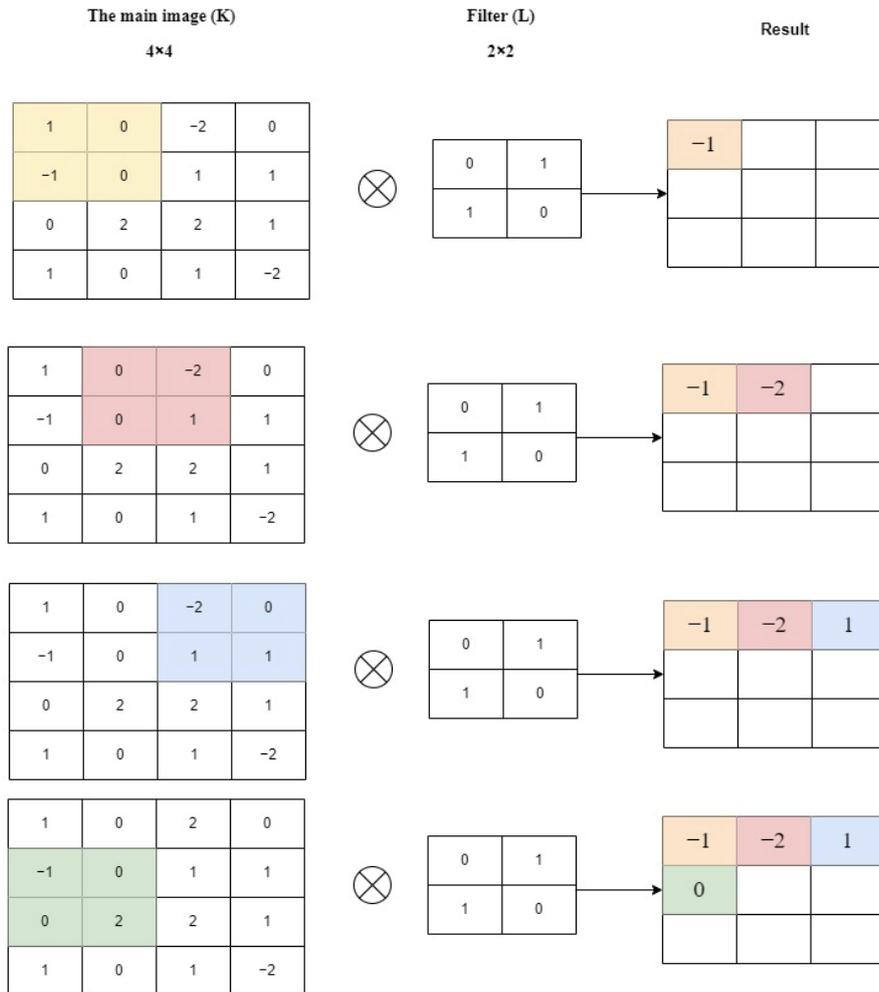


Figure 3.3: Visual representation of the calculations involved in the convolutional layer. Original source: [Taye 2023](#)



Figure 3.4: Effects of different convolutional matrices on the initial image (top left). Each region of the figure displays a kernel and the resulting image is obtained through the convolution of the starting image with the corresponding kernel. The matrix in the center-left corresponds to a *blur* kernel, intended to diminish differences in adjacent pixel values. Conversely, the bottom-left kernel represents a *sharpen* kernel, enhancing disparities in adjacent pixel values. The top-right kernel highlights significant variations in pixel values. In the center and bottom-right, two types of *Sobel* kernels are depicted, showcasing variations in pixel values along specific directions.

Usually, convolutional layers and activation functions are stacked in series, which allows the network to capture abstract features. Usually, the pooling layers are applied at the end of the model. The final product of the feature extractor is a flattened feature vector that serves as the input for a standard MLP, which ultimately makes the predictions.

3.3 k-Nearest Neighbors Algorithm

The k-nearest neighbor algorithm (*kNN*) is a non-parametric supervised learning classifier. Its primary function is to classify a given data point by determining its group or category. This is based on the assumption that similar data points tend to cluster together. *kNN* is a widely recognized and used algorithm known for its simplicity and accuracy. The *kNN* approach was first formulated by [Fix and Hodges \(1951\)](#) and subsequently expanded by [Cover and Hart 1967](#). The *kNN* algorithm has numerous applications, such as pattern recognition, data mining, recommendation systems, intrusion detection, and financial predictions.

The exact search problem that kNN has to solve is:

- Given a set of points E in a d -dimensional space, identify the k points in E with the smallest distance to any query point q .

The problem is fully specified defining the metric distance in the particular d -dimensional space E . Usually, the metric is an l_s norm:

$$\|p - q\|_s \quad \text{where} \quad \|x\| = \left(\sum_{i=1}^d |x_i|^s \right)^{1/s}. \quad (3.8)$$

There are two main ways to address this search problem:

1. **Brute force:** When given a query q , the algorithm calculates the distance between q and every point in E , then selects the k points with minimum distances. The computational complexity of this scan is $O(kdn)$, where n is the number of points in the dataset.

2. **Data structures:** The dataset is divided in data-structures. So, when given a query, the algorithm can speed up the computation time.

Typically, tree structures (a type of graph) are used as the underlying data structure. Thus, let us briefly introduce the theory behind graphs and tree structure.

The graph G consists in a set of points, called *nodes* $V(G)$, together with a set of *edges* $E(G)$ connecting the *nodes*. This structure represents pair-wise relations between objects (see Fig 3.5).

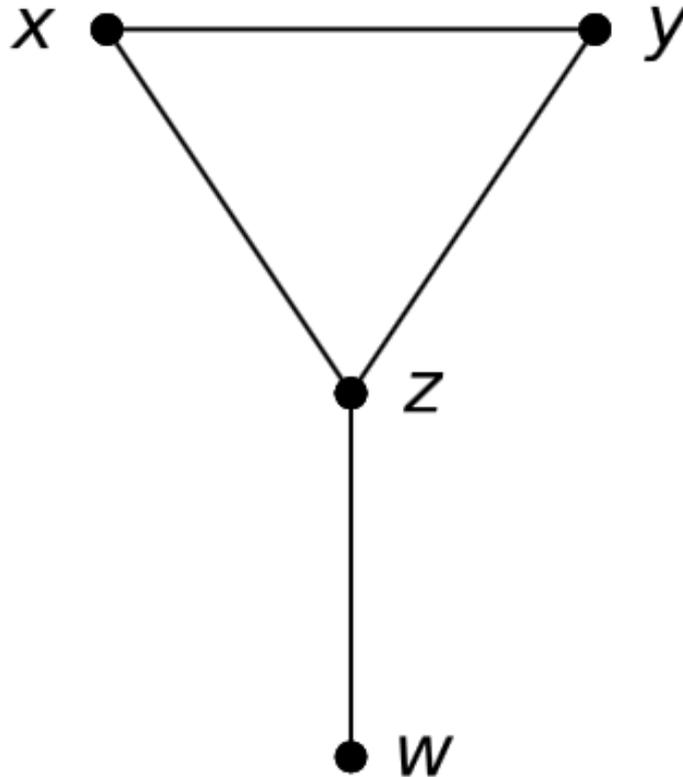


Figure 3.5: Visual representation of the graph $G(V, E)$ where the node set is $V(G) = \{x, y, z, w\}$ and the edge set is $E(G) = \{x, y\}, \{x, z\}, \{y, z\}, \{z, w\}$

The graphs are defined by many properties, for our purposes we define the following:

- **Directional Graphs:** The edges in directional graphs possess a specific orientation. The relationships connecting the nodes are inherently asymmetric, differently from Fig. 3.5.
- **Connected Graph:** A graph G is considered connected if, for any distinct vertices $x, y \in V(G)$, there exists a path from x to y . Like the graph in Fig. 3.5.
- **Cycle:** In graph theory, a cycle is discerned within a graph when a non-empty path is identified, characterized by equivalent initial and terminal vertices. The graph in Fig. 3.5 does contain a cycle between the nodes x, y, z .

A graph G is called *tree* when it is connected and does not contain any cycle. Usually, directional trees are used to represent hierarchical structure (Fig. 3.6) and they have the following structure:

1. One node is particular and named the *root*. Typically, it is represented at the top.

2. Each non-root node c is connected to another node p through an edge, referred to as *parent* node. Similarly, c is considered a *child* node of p .
3. The nodes which do not have any child are called *leaves* or external nodes.
4. The depth of a node is defined as is the length of the path to its root. The height of a tree is defined as the length of the longest downward path to a leaf from the root.

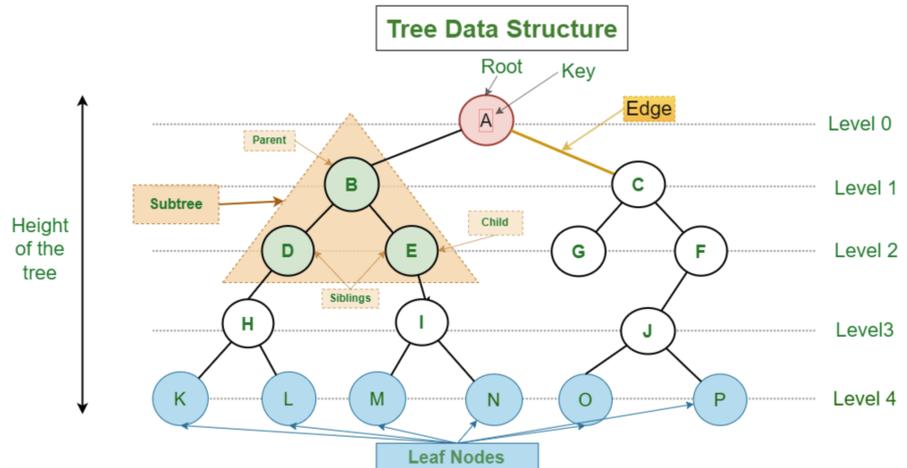


Figure 3.6: Visual representation of a tree data structure

Tree data structures are valuable for organizing datasets hierarchically, effectively reducing the computational time required to retrieve points with specific properties within the dataset.

For the retrieval of k nearest neighbors of a given query q , two tree structures are employed:

- **K-D Tree:** It is a tree data-structure invented by [Bentley 1975](#). Given a set of n points in a d -dimensional space, the K-D tree is built recursively. First, it finds the median value of the i -th coordinate of the points, with $i = 1$ initially. This value, x , is stored, and the dataset is divided into two E_R and E_L , with E_R containing all the points with their i -th coordinate greater than x and E_L being the complementary set of points. The process is repeated recursively, replacing i by $i + 1$. Every non-leaf node can be conceptualized as implicitly creating a splitting hyperplane, thereby partitioning the space into two distinct regions referred to as half-spaces. Points situated to the left of this hyperplane are represented by the left subtree of the respective node, while points located to the right of the hyperplane are characterized by the right subtree. The algorithm ends when the set of points at a certain node becomes 0, see Fig. 3.7. The resulting structure has a height of $\log n$ and n leaves, the time complexity in building the structure is of $O(dn \log n)$.

To retrieve the nearest neighbor of a given query q , the algorithm operates as follows: At each step, it calculates and records the distance R to the nearest point encountered so far. Initially, $R = \infty$. In a leaf node, which contains a point p , the algorithm checks if $d = \|p - q\|_s$ is less than R . If this is true, then R is set to d , and p is saved as the closest point candidate. Then, the algorithm starts to travel the tree backwards. At each point, it computes the distance from the current node, if lower than the current best, it becomes the new current best. In addition, it determines whether the distance between the splitting coordinate of the search point and the current node is less than the overall distance from the search point to the current best. If true, there could be nearer points on the other side of the splitting and the algorithm must go down that branch. Conversely, if false, it can continue the travel backward. The algorithm stops when reaches the root.

Note that this algorithm can easily be modified to find k -nearest neighbors. In this case, the elimination of a branch occurs only when k points have been identified.

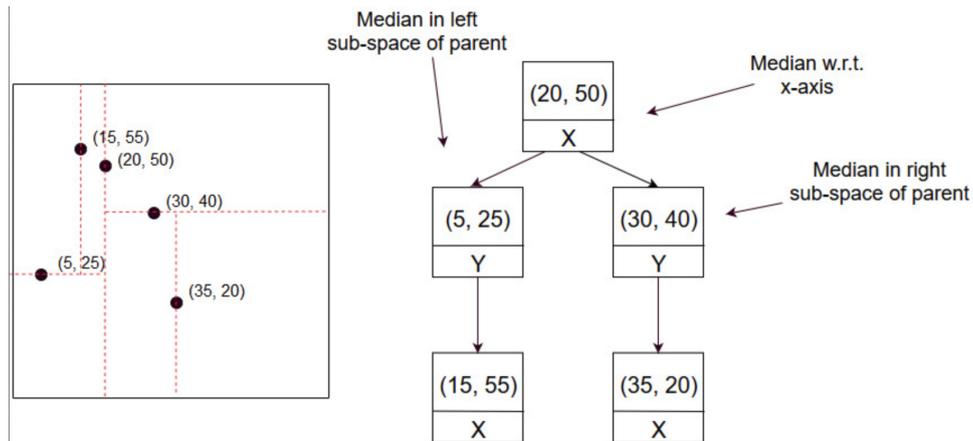


Figure 3.7: Visual representation of a K-D tree. The left side illustrates the dataset partition as previously explained. On the right, the equivalent structure is depicted using a tree structure.

- **Ball-tree:** A Ball-tree is a binary tree in which each node has a “ball” associated with it, such that the ball of an inner node is the smallest ball that contains the “balls” of its children. For *ball*, we refer to a hyper-sphere in the n -dimensional space \mathcal{R}^n . The dataset is partitioned in two, along the dimension with the biggest spread of the data.

During the search problem, the algorithm manages a max-first priority queue, denoted as Q , representing the k closest points to date. To update Q , it acts as follows: If the distance from the query point q to the current node A exceeds the distance to the furthest point in Q , disregard A and return Q . If A constitutes a leaf node, scan each point in A and adjust the nearest-neighbor queue accordingly. In the case where A is an internal node, the algorithm act recursively on the two children of A , prioritizing the search on the child whose center is closer to the query point q . After each of these recursive calls, the algorithm updates Q

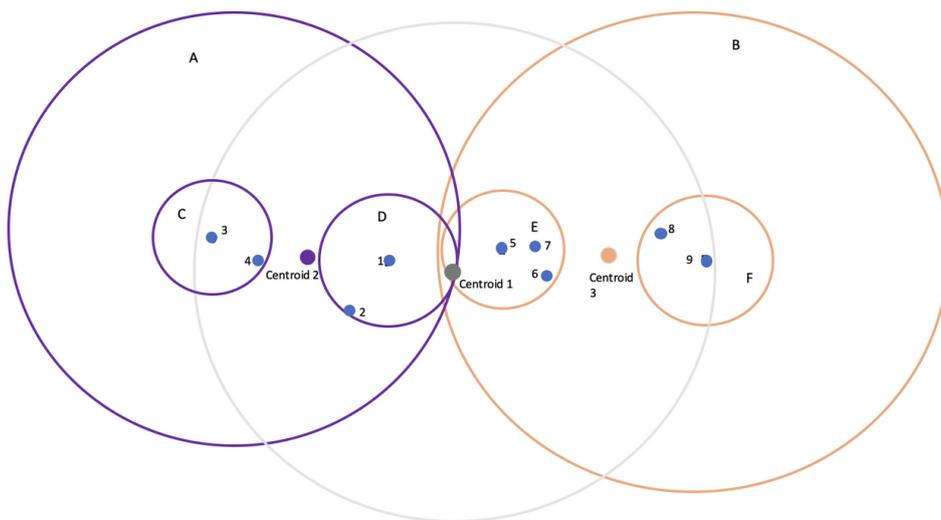


Figure 3.8: Schematic representation of the dataset partition using the ball-tree algorithm

The key contrast between the two structures is their effectiveness in different dimensional spaces.

The first structure performs better in low-dimensional spaces, while the second is more efficient in high-dimensional spaces. This distinction stems from the distinct geometric partitioning methods employed by each structure on the dataset.

3.4 ULISSE

ULISSE (Doorenbos et al. 2022) employs EfficientNet-b0 (Tan and Le 2019), a particular CNN, as the feature extractor by excluding its final layers. The network is trained on the well-known ImageNet dataset (Deng et al. 2009), which consists of 1.3 million common images (such as dogs, cats, bridges and so on) sorted into 1000 classes. During feature extraction, the CNN transforms each image into a vector of 1280 features, providing an image representation in a lower-dimensional space relative to the original. As previously said, CNNs learn, through training, to prioritize the most essential features based on its training dataset. Despite ImageNet being different from an astronomical dataset, we expect that ULISSE can be useful even when applied to other types of images. Even if the performance is not ideal, the features extracted should still be relevant due to their abundance. This leads us to the primary advantage of ULISSE: it operates without the need for a task-specific dataset. This is particularly important for gravitational lensed objects, as a real dataset for such cases is currently unavailable.

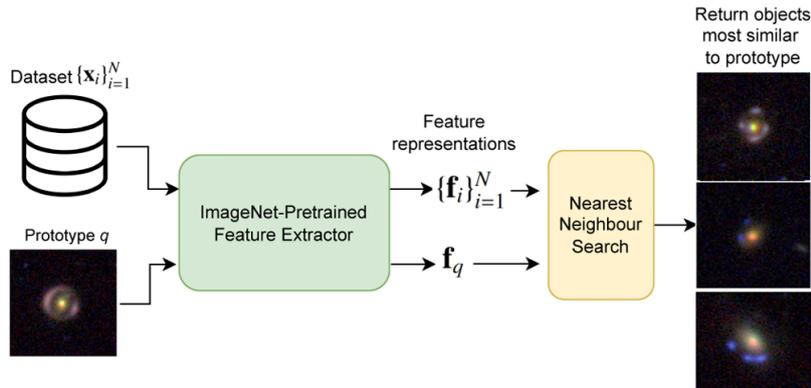


Figure 3.9: Overview of ULISSE. Original source : Doorenbos et al. 2022.

ULISSE is a useful tool for identifying the most similar objects to a given query within a general image dataset. It breaks down both the query object and the dataset into features to find the most similar candidates. To achieve this, it conducts a kNN similarity search in the feature space of the initial dataset. A list of candidates is generated by the objects closest to a given prototype in the feature space. Specifically, when provided with a prototype image x_q , the goal is to retrieve its closest neighbors from a dataset $f(x_i)_{i=1}^N$. For an image, we denote its feature representation as $f(x_i) = f_i$. To find the nearest neighbors, we use their distance in this trained feature space. As a result, we identify those objects that minimize $d(x_q, x_i)$. The use of acceleration structures such as tree structures enables efficient computation of nearest neighbor searches (see more in Sec 3.8). Consequently, these lookalikes are extremely fast to find after their initial construction. If there is no validation data available, ULISSE can be utilized by providing a list of the n nearest objects in the dataset, measured in feature space, rather than astronomical distance. The user can adjust n according to their needs. If a validation set is present, we can determine a performance threshold based on the number of nearest neighbors.

Chapter 4

Data

Collecting a large dataset is crucial for assessing the performance of ULISSE in the search of new gravitational lenses. Because we need to distinguish gravitational lenses by galaxies, the dataset should encompass both instances of non-lensed galaxies (negatives) and galaxies exhibiting strong gravitational lensing effects (positives). For the former category, we have used real three-bands Kilo-Degree Survey DR4 galaxies, whereas the latter is constructed using a hybrid approach by superimposing simulated images to real KiDS galaxies. This approach is adopted due to the limited number of confirmed gravitational lenses. Specifically, those derived from comprehensive searches such as the Sloan Lens ACS Survey (SLACS, [Bolton et al. 2006](#); [Shu et al. 2017](#)), are less than one hundred, and in total the number of known lenses reaches few hundreds elements. This quantity proves insufficient for conducting a robust statistical analysis of ULISSE’s performances.

The introductory section of this chapter (Sec. 4.1) is dedicated to presenting the Kilo-Degree Survey. We delineate its observational strategy following a compact overview of its principal scientific objectives and secondary science cases. The following section (Sec. 4.2) focuses on the details of the Dataset and discusses the simulation technique used to generate lensed galaxies.

4.1 The Kilo Degree Survey (KiDS)

The Kilo-Degree Survey (KiDS, [de Jong et al. 2013](#)) is one of three public ESO surveys using the 256-megapixel camera OmegaCAM ([Kuijken 2011](#)) attached to the Cassegrain focus of the ESO VLT Survey Telescope ([Schipani et al. 2012](#)) at Paranal Observatory in Chile. The OmegaCAM is composed by 32 CCD detectors that cover a 1 square degree field of view with a pixel size of 0.21 arcseconds. The VST is a 2.6-meter telescope with active control of both the primary and secondary mirrors. As a result of this design choice, the camera-telescope combination is well-suited for capturing clear and uniform images over a wide field of view with, typically, sub-arcsecond images.

KiDS is designed to study the effect of weak gravitational lensing on galaxy shape, due to structures along the LoS. Combining this measurement with photometric redshift estimates, it is possible to obtain a weak gravitational lensing tomography. By tracking the correlation in galaxy shapes across different redshifts, this method, namely *cosmic shear*, traces the evolution of large-scale structures (for more information see [Mellier 1999](#); [Kilbinger 2015](#)). This provides valuable insights into the dark sector of the universe, encompassing dark energy and dark matter. Ultimately, the technique coupled with the ability to measure the geometry of the Universe could, potentially, allow us to distinguish between dark energy and modified theory of gravity (see more in [Hu 1998](#); [Capozziello and de Laurentis 2011](#)).

In order to meet this goal, KiDS observe the extragalactic sky, targeting galaxy population out to redshift ~ 1 , in four bands (*ugri*). The *r*-band is used to measure the shape of the galaxies, thus the best seeing conditions are reserved for it. To obtain photometric redshift

estimates, the r -band is complemented by the g,u and i -bands images taken in progressively worse seeing condition. The i -band is the only filter observed during bright time condition. The KiDS observing strategy is summarized in Table 4.1.

Moreover, KiDS observations are further complemented by VIKING’s (Edge et al. 2013) data, which is obtained on the same field of the sky but in other five bands (z , Y , J , H and K_s). The end result is a 9-band optical-infrared dataset. KiDS, since its inception, has targeted around 1350 squares degree of extragalactic sky divided in two fields to provide year-round observability, see Fig. 4.1.

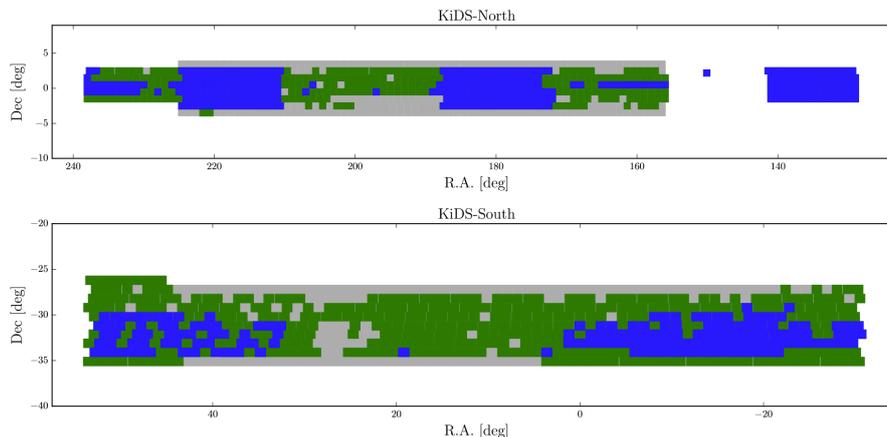


Figure 4.1: Illustration of the spatial distribution of survey tiles unveiled in KiDS-ESO-DR4. The tiles marked in green represent their initial release, whereas those marked in blue were included in prior data releases but underwent reprocessing for DR4. The overall KiDS+VIKING expanse is portrayed in gray, with the upper segment representing KiDS-North and the lower segment representing KiDS-South. Significantly, the single point at $RA = 150$ degrees is centrally positioned on the COSMOS/CFHTLS D2 field. Original source: Kuijken et al. 2019.

Table 4.1: KiDS observing strategy

Filter	Max. lunar illumination	Min. moon distance [deg]	Max. seeing [arcsec]	Max. airmass	Sky transp.	Dithers	Total Exp. time [s]
u	0.4	90	1.1	1.2	CLEAR	4	1000
g	0.4	80	0.9	1.6	CLEAR	5	900
r	0.4	60	0.8	1.3	CLEAR	5	1800
i	1.0	60	1.1	2.0	CLEAR	5	1200

Note: Lunar illumination is defined as the fraction of the lunar disk that is illuminated at local (Chile) civil midnight, here 1.0 is fully illuminated. Moon distance specifies the angular distance between the astronomical target and the Moon. Max. seeing is the maximum full width at half-maximum of the point spread function. Airmass is a measure of the amount of air along the line of sight. CLEAR Sky transparency means less than 10% of the sky (above 30 degrees elevation) covered in clouds. More information of definition of observation condition at <https://www.eso.org/sci/observing/phase2/ObsConditions.html>

Beyond its primary scientific aim (Hildebrandt et al. 2017a,b; Shan et al. 2018; Asgari et al. 2019), the KiDS dataset supports diverse secondary science cases. These include the study of galaxy evolution (Roy et al. 2018; Tortora et al. 2018a), environment (Brouwer et al. 2016), bias (Dvornik et al. 2017) and galaxy-halo connections (van Uitert et al. 2017); solar system objects (Mahlke et al. 2018), photometric estimates through machine learning methods (Cavuoti et al. 2017; Amaro et al. 2019) and the search for rare objects such as strong lenses (Petrillo et al.

2019a; Li et al. 2020; 2021), ultra-compact massive galaxies (Tortora et al. 2018b), and quasars (Spiniello et al. 2018).

4.2 The Dataset

To adequately evaluate the algorithm, a robust *knowledge base* is indispensable, necessitating a dataset of pre-labeled images with a known ground truth. For our study, the knowledge base, for the positive data, is constructed using 16,000 (29% of the total dataset) simulated gravitational arcs provided by Li et al. 2021. Whereas, the negative dataset (71% of the total) comprises 25,000 Luminous Red Galaxies (LRGs) from the KiDS DR4 survey, and approximately 15,000 KiDS DR4 galaxies with magnitudes brighter than $r = 21$, collectively referred to as the bright galaxy sample (see Fig. 4.2). It is noteworthy that the LRGs form a subset of the bright galaxy sample.

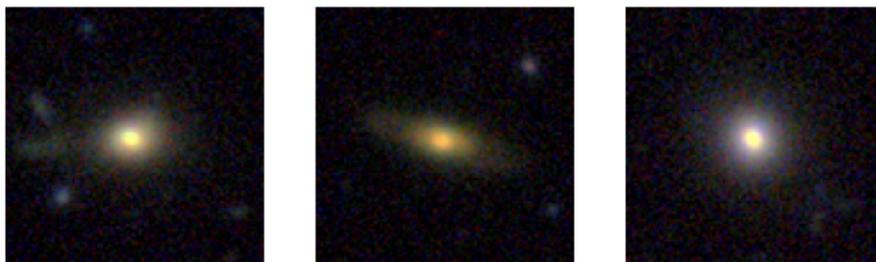


Figure 4.2: Three examples of the bright galaxy sample

4.2.1 Luminous red galaxy sample

LRGs are relatively massive galaxies ($10^{11} - 10^{12} M_{\odot}$), elliptical systems composed of old stars. Being very massive, are more likely to exhibit lensing features compared to other classes of galaxies ($\sim 80\%$ of the lensing population, Turner et al. 1984; Fukugita et al. 1992; Kochanek 1996; Chae 2003; Oguri 2006; Möller et al. 2007). The low- z ($z < 0.4$) selection criteria were defined by Eisenstein et al. (2001) and adapted by Petrillo et al. (2017) to include fainter and bluer sources:

$$\begin{aligned} r_{auto} &< 14 + c_{par}/0.3, \\ |c_{perp}| &< 0.2, \end{aligned}$$

where,

$$\begin{aligned} c_{perp} &= (r - i) - (g - r)/4.0 + 0.18, \\ c_{par} &= 0.7(g - r) + 1.2[(r - i) - 0.18]. \end{aligned}$$

Here, r_{auto} represents the r-band Kron-like magnitude¹ (Kron 1980), specifically the *MAG_AUTO* value acquired using SExtractor (Bertin and Arnouts 1996). The photometric redshift z is determined through the BPZ⁽²⁾ code (Benítez 2000). Colors $(g - r)$ and $(r - i)$ are computed

¹To calculate the ‘‘Kron Magnitude,’’ it is necessary to define the ‘‘Kron Radii’’ as follows:

$$R_1(R) = \frac{2\pi \int_0^R I(x)x^2 dx}{2\pi \int_0^R I(x)x dx}. \quad (4.1)$$

A circular aperture with a radius twice R_1 , where R_1 is determined by integrating up to a radius R that represents 1% of the sky flux, encompasses more than 90% of an object’s total light. SExtractor computes its magnitudes using apertures that are 2.5 times R_1 .

²<http://www.stsci.edu/~dcoe/BPZ/>

within Gaussian-weighted apertures, as outlined in [Kuijken et al. 2015](#). These apertures are customized on a per-source and per-image basis, ensuring that they yield flux estimates independent of the varying observational conditions and bands ([Kuijken et al. 2019](#); [Petrillo et al. 2019a](#)). Furthermore, a cutout of 101×101 pixels has been extracted, centered on each of these galaxies, corresponding to 20×20 arcsec². These cutouts are sufficiently large, encompassing galaxy-sized and multiple image regions. They also provide a view of the environment surrounding the identified lens candidates. This cutout size translates to $90 \text{ kpc} \times 90 \text{ kpc}$ at $z = 0.3$ or $120 \text{ kpc} \times 120 \text{ kpc}$ at $z = 0.5$. In addition to these criteria, the LRG dataset underwent visual examination to exclude galaxies that might exhibit arc-like characteristics around them, resembling a gravitational lensing occurrence. Additionally, galaxies with suboptimal image quality or influenced by stellar diffraction spikes or reflection halos were also screened out. RGB images were constructed using the HumVI algorithm developed by [Marshall et al. 2015](#).



Figure 4.3: Three examples of LRG sample

4.2.2 Simulation

The simulation process starts by modeling the mass profile of the lens as a SIE ([Kormann et al. 1994](#), see also [Sec 2.3.4](#)), and the light profile of the source with a Sérsic profile ([Sérsic 1963](#); [Sersic 1968](#), see also [Appendix B](#)). Through the lens equation ([Eq. 2.6](#)) it is possible to generate gravitational arcs which are then superimposed on a random galaxy from the LRG sample. Finally, the image obtained is convolved with a Moffat PSF.

Simulating the deflector

The simulation procedure starts by modeling the mass distribution of the lens through a Singular Isothermal Ellipsoid model with external shear to account for the effect of different environments ([Keeton et al. 1997](#), [Sec 2.2](#)). In addition, the mass distribution is perturbed by an additional Gaussian Random Field (GRF) fluctuation with a power law spectrum. The power-law exponent is fixed to -6 , which is a good, first-order approximation, of substructures within the lens plane in the Λ CDM cosmological model ([Hezaveh et al. 2013](#); [Chatterjee and Koopmans 2018](#)). The variances for determining the amplitude are drawn from a logarithmic distribution between $10^{-4} - 10^{-1}$ about mean zero in the units of the square of the lensing potential ([Petrillo et al. 2019a](#); [Li et al. 2021](#)). This results in a lens mass profile that is not a perfect SIE. The parameters are sampled following the distribution in [Table 4.2](#).

Table 4.2: Parameters range and distribution for the simulated mass distribution of the lens (SIE)

Parameter	Range	Units	Distribution
Einstein radius	1.0 – 5.0	arcsec	exponential
Axis ratio	0.4 – 1.0		uniform
Position angle	0 – 180	degree	uniform
External shear	0 – 0.1		uniform
Angle of external shear	0 – 180	degree	uniform

The upper limit of 5 arcsec for the Einstein radius aims to include typical Einstein radii for strong galaxy–galaxy and group–galaxy lenses (Koopmans et al. 2009; Foëx et al. 2013; Verdugo et al. 2014). The lower limit is similar to the FWHM of the r-band KiDS PSF. The axis ratio > 0.4 of the lens excludes spiral and very elliptical galaxies (Binney et al. 2000).

Simulating the source

Once the mass distribution of the deflector is simulated, a source parametrized with a Sérsic profile is added. The parameters of the model (Sérsic index, effective radius, axis ratio and position angle) are sampled from the ranges and distribution reported in Table 4.3

Table 4.3: Parameters for the Sérsic source

Parameter	Range	Units	Distribution
Effective radius (R_{eff})	0.1 – 0.5	arcsec	Normal ($\mu = 0.2, \sigma = 0.3$)
Axis ratio (q)	0.3 – 1.0	–	Uniform
Position angle	0 – 180	degrees	Uniform
Sérsic index (n)	0.3 – 5.0	–	Uniform

The sources’ position are chosen uniformly within the radial distance of the tangential caustics plus one effective radius of the source Sérsic profile. Therefore, the training set is mostly composed of high-magnification rings, arcs, quads, folds, and cusps rather than doubles (Schneider et al. 1992). For each source, a realistic color is simulated to create images in g, r, i -bands. To produce realistic 3-bands images, the magnitudes are extracted from the Rubin/LSST mock galaxies catalog (Connolly et al. 2010). The catalog has photometric depth of $r \sim 28$ and covers the redshift range $0 < z < 6$. It was generated from the Millennium Simulation (Springel et al. 2005), with superimposed galaxies based on a semi-analytical model for galaxy evolution (De Lucia et al. 2006) which include gas cooling, star formation and supernova/AGN feedback, to reproduce observed colors, luminosities and clustering of galaxies. Li et al. (2021) selected ~ 2600 of these mock galaxies at redshift between 0.8 and 3 and with r -band AB magnitude between 21 and 25. The magnitude of the sources are randomly selected from this color library. Finally, Li et al. (2021) perturbed the magnitudes in all three bands by randomly adding a value of ± 0.1 to account for some scatter around the nominal Rubin/LSST mock colors, similar to what has been done in Petrillo et al. 2019a.

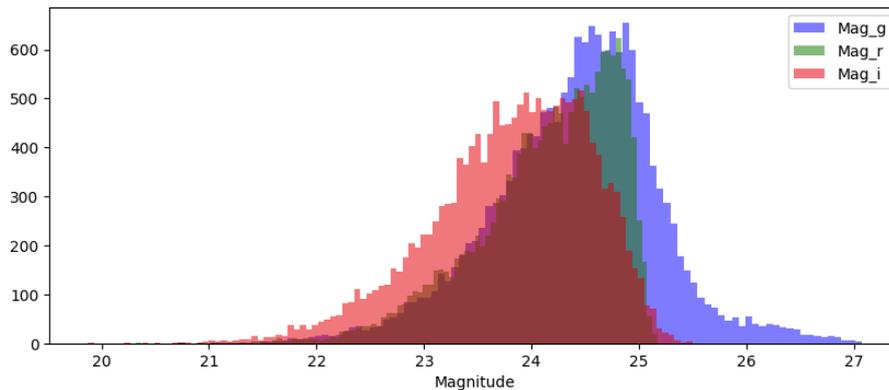


Figure 4.4: Three-band magnitudes (g, r, i) of the Sérsic source.

At source redshifts of $z > 0.5$, smaller sizes and smaller Sérsic indices are found with respect to the local universe, and the fraction of spiral galaxies (with $n = 2 - 3$) increases (Trujillo et al. 2007; Chevance et al. 2012). The uniform sampling has been employed for all the parameters except the effective radius, which follows a Gaussian distribution to mimic the small size of the high- z galaxies.

Point Spread Function

The simulated arcs/point images underwent convolution with a simulated Point Spread Function (PSF). The simulated PSF employed in the study by Li et al. (2021) follows a Moffat profile (for more information see Appendix C). The range and distribution of the parameters were set accordingly to the seeing distribution of KiDS (Kuijken et al. 2019) and previous experience in modelling the PSF parameters (e.g, La Barbera et al. 2008; Roy et al. 2018).

Table 4.4: PSF parameters

Parameter	Range	Units	Distribution
FWHM-g	0.60 - 1.2	-	Normal ($\mu = 0.85, \sigma = 0.1$)
FWHM-r	0.50 - 0.90	-	Normal ($\mu = 0.7, \sigma = 0.05$)
FWHM-i	0.55 - 1.2	-	Normal ($\mu = 0.80, \sigma = 0.1$)
<i>Additional Parameters</i>			
β	2.20	-	Fixed
Axis ratio	0.98 - 1.02	-	Uniform
Position angle	0 - 180	-	Uniform

PSF variation leads to a large variance in the sharpness of the lensing features, this impacts mostly g and i -bands images (see Li et al. 2021).

Final Normalization

In choosing, which LRGs to superimpose, it is required that the ratio of the peak of the r -band surface brightness of the arcs α and the LRGs β to satisfy the condition $\alpha/\beta \geq 0.05$, or the local signal-to-noise ratio of the surface brightness peak of the arcs (3×3 pixels with the peak as center) to be $SNR > 5$. The results are images like the ones shown in Fig. 4.5.

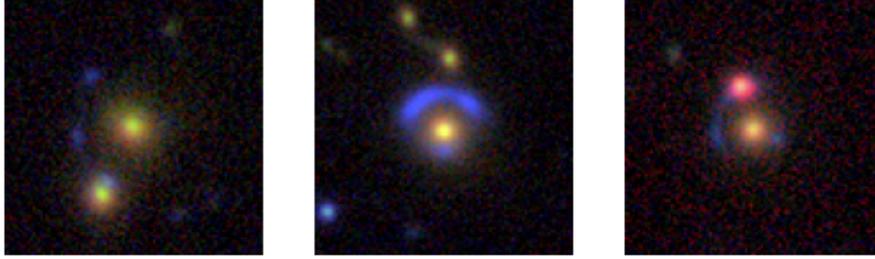


Figure 4.5: RGB simulated images of gravitationally lensed objects. The source at the center of the image is an LRG, which has been superimposed on the bluer simulated gravitational arc.

The final distributions of the parameters after the normalization are shown in Fig. 4.6.

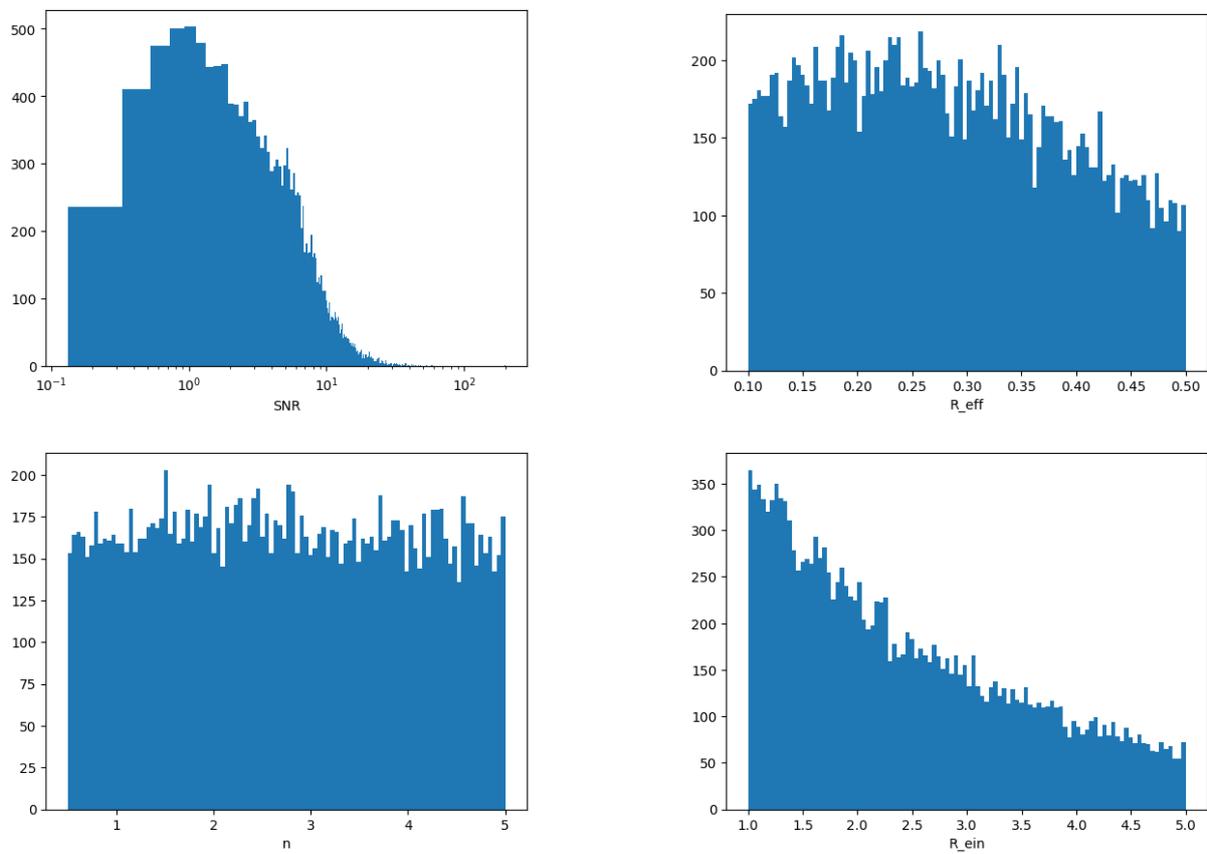


Figure 4.6: Final distributions of the SNR of the gravitational arc, the effective radius of the source, Sérsic index of the source, and Einstein radius of the lens.

Chapter 5

Results

ULISSE operates by taking an image, denoted as the “query”, and providing a set of k similar images, referred to as “candidates” or k -nearest neighbors (as detailed in Chapter 3). The search for similarities is conducted within a dataset, that excludes the query itself. To ensure a robust statistical analysis, multiple queries must be used, and ULISSE performs similarity searches on an extensive dataset. The performance of ULISSE on a given query is defined through what we named the *accuracy* of ULISSE. It is determined by the ratio of candidates having the same label (positive or negative) of the query to the total number of candidates.

Our initial focus involved the optimization of different parameters associated with the dataset and ULISSE. This encompasses parameters such as the number k of candidates, image band and scale, distance metrics, and cutouts. Additionally, we explored the relation between the algorithm’s performance and the simulation parameters of the query.

Finally, the performance of ULISSE has been further investigated in a quasi-realistic case, in which real lens candidates have been used as queries and the similarity-search has been performed on the (half-simulated) Dataset described in Sec 4.2. Also, the reverse has been performed. The aim is to evaluate the applicability of ULISSE in real case scenarios and in doing that also performing the pruning of the various parameters involved.

5.1 Tuning of the parameters

Firstly, we established the parameters that maximize ULISSE’s performance. The Dataset was divided into two mutually exclusive subsets, a small dataset of query images and a complementary set. The query set acts as the input for ULISSE which then returns a list of candidates found on the complementary set. Two considerations drove this decision. Firstly, despite ULISSE being fast computationally, the high number of parameters means that the number of queries should be low. Secondly, to mimic a realistic scenario, the complementary dataset must be sufficiently extensive. For these reasons, we initially elected to use 100 queries, containing 50 positive and 50 negative examples. The queries were randomly selected from the entire Dataset to (try to) ensure a representative sample.

With RGB queries, we obtained that the average accuracy of ULISSE on the positive queries is of 65,9% within its first 100 neighbors. This means, that if we feed ULISSE with a three-band lensed object, within the list of the 100 candidates, found by ULISSE, we expect to find 65,9 positive images. A good reference value to compare this accuracy with is the random classifier accuracy (in this case a model that returns a list of 100 candidates completely random) which is 29%.

The first thing we did was to search the best value of k , keeping in mind that a low number of neighbors is preferred to quickly inspect the list of candidates (in a real case situation). Thus, we plotted the average accuracy against the number of nearest neighbors, in the case of positive queries and negative queries (Fig. 5.3 and 5.4).

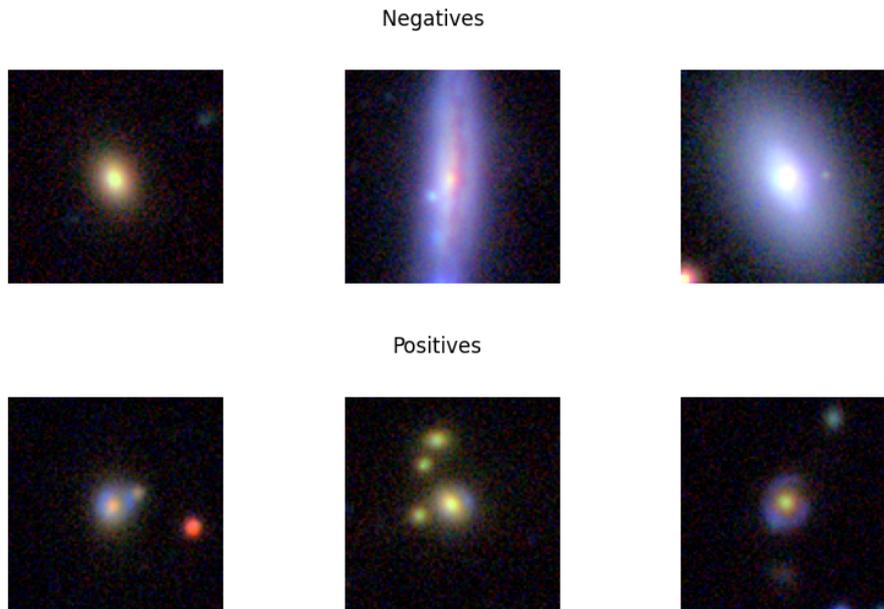


Figure 5.1: Examples of the Dataset. The RGB images on top shows real galaxies (negatives) whereas the RGB images in the bottom show lensed simulated galaxies (positives). The images are (g, r, i) RGB images ($20'' \times 20''$)

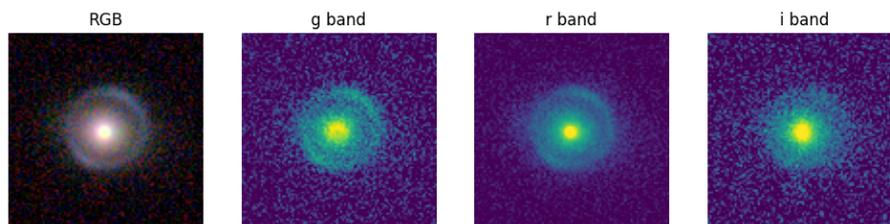


Figure 5.2: Representation of the same lensed galaxy in RGB and individual bands. The RGB image is presented on the left, followed by the g -band, r -band, and i -band.

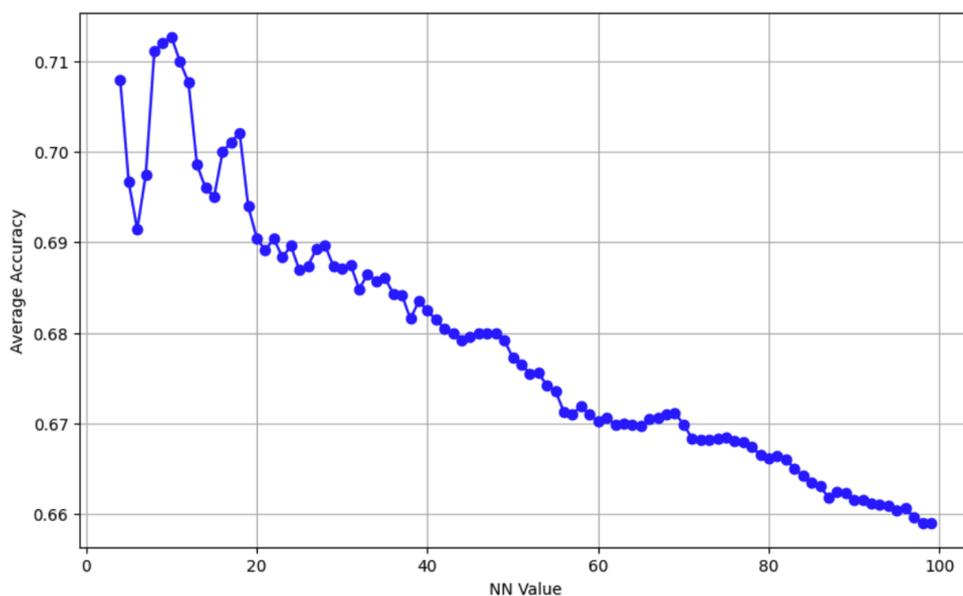


Figure 5.3: Average accuracy of positive queries. The plot represent the average percentage of positive (y-axis) examples at different nearest neighbors (x-axis) within the similars of positive queries .

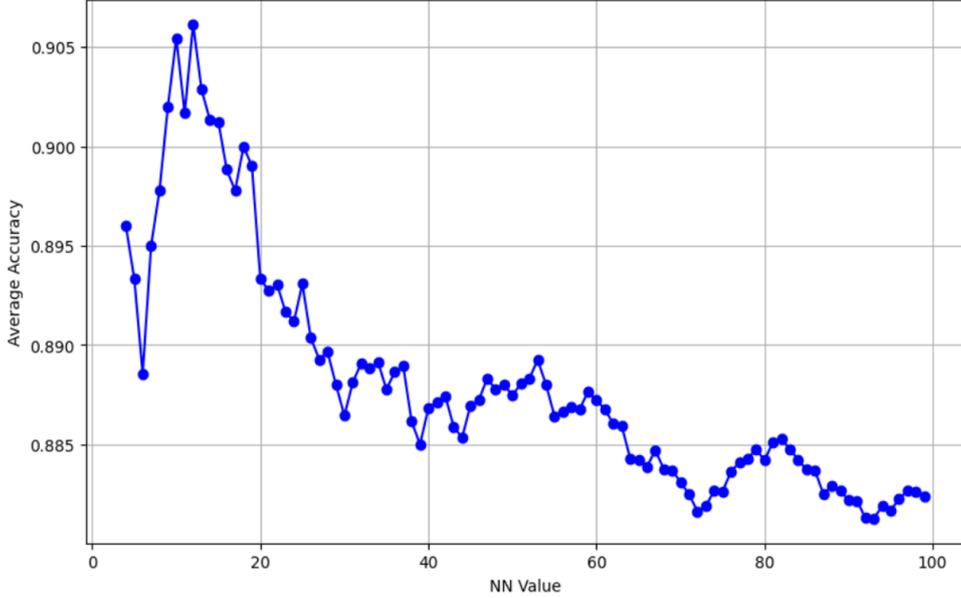


Figure 5.4: Average accuracy of negative queries. The plot represent the average percentage of negative (y-axis) examples at different nearest neighbors (x-axis) within the similars of negative queries.

The best accuracy for positive queries was observed at 71%, coinciding with the 10 nearest neighbors mark (see Fig. 5.3). Luckily, the negative queries exhibited a similar peak in accuracy around the same value of k . Therefore, for uniformity and optimal performance, we have decided to proceed with selecting only the first 10 nearest neighbors for subsequent analysis.

The next test conducted considered the performance of ULISSE on the same set of images, but only with one of the three spectral bands (see Fig 5.2). Therefore, we computed the average accuracies of the queries on the first 10 nearest neighbors. Specifically, for the i band, the accuracy of positive queries is of 40%, a notable difference from our previously established reference accuracy of 71%. Conversely, in the cases of the r and g spectral bands, we obtained accuracies of 58% and 69%, respectively.

We also explored the impact of image scale on ULISSE’s performance. Adopting a logarithmic scale¹, the mean accuracy for the three spectral bands positives queries is 63%, surpassing the individual band values which fell below 40% for each band. Employing the hyperbolic sine scale² yielded an accuracy of 68%, a modest improvement compared to the logarithmic case, though still falling short of the reference accuracy.

We also experimented with cutout size. It is important to acknowledge that this approach, while simplifying ULISSE computations, introduces the potential for both the elimination of irrelevant features but also of important features, like the environment around the lensed object. Moreover, the outcomes of this analysis are inherently dependent upon the characteristics of the dataset under consideration. The obtained values for this analysis are as follows: for a square cutout of 80 pixels, an average accuracy of 71% was obtained. Cutouts of sizes 65 and 50 exhibited identical average accuracies of 74%. The best accuracy, however, was attained with a cutout size of 40, yielding an accuracy of 75%. Beyond this point, the accuracies experienced a sharp decline, imputable to the removal of important features from the gravitational arcs. In fact, about 10% of the simulated lenses exhibit an Einstein Radius exceeding 4 arcseconds. Given that the pixel size of these images is 0.2, this implies that the arc of these objects will be cropped.

¹The scaling formula applied to the values of the pixels is $y = \log(x + 1)$

²The scaling formula applied to the values of the pixels is $y = \operatorname{arcsinh}(x)$

The distances calculated by ULISSE in the feature space were determined using the Euclidean metric. However, the choice of this metric is not inherently justified as the optimal one. The well-known challenge associated with the Euclidean metric in high-dimensional spaces, commonly referred to as the “curse of dimensionality” (Bellman 1957), prompted us to investigate alternative distance metrics. Therefore, we conducted tests using the cosine similarity³ and Manhattan distance⁴. The resulting accuracies were similar to those obtained with the Euclidean metric. Consequently, for the subsequent phases of our study, we maintained the use of the Euclidean distance metric.

5.2 Statistics

Having identified the optimal settings for ULISSE, which is three-bands linear images, and employing a 40-pixel cutout, we aimed to improve the statistical significance of the average accuracy measure. Instead of dividing the Dataset in two, we adopted a different approach by computing the accuracy of each individual image, in the role of the query, against the entirety of the Dataset without the query. In particular, the average for the positives is computed over a group of 16500 examples whilst for the negatives for ~ 40000 images. The accuracy obtained for the positive queries is of 59% while for the negatives is 87%.

Another approach to find lensed objects, having a numerous set of positive queries, could be to look at the most frequent images within the list of candidates (of the positive queries) that ULISSE returns. For instance, in this study, the fifty most frequent images comprehend 47 positives (94%) and 3 negative images, see Fig. 5.6. The negatives have been encased in red rectangles to be distinguished, they are at position 17th, 33rd, and 47th for frequency.

³The cosine similarity is defined as the Euclidean dot products:

$$A \cdot B = |A||B| \cos \theta, \quad (5.1)$$

where $\cos \theta$ is the angle between the vector A and the vector B. Basically, the cosine similarity measures the difference in orientation between two vectors.

⁴The Manhattan distance between two points $x = (x_1, x_2, \dots, x_n)$ and $y = (y_1, y_2, \dots, y_n)$ is defined as:

$$d_M(x, y) = \sum_{i=1}^n |p_i - q_i| \quad (5.2)$$

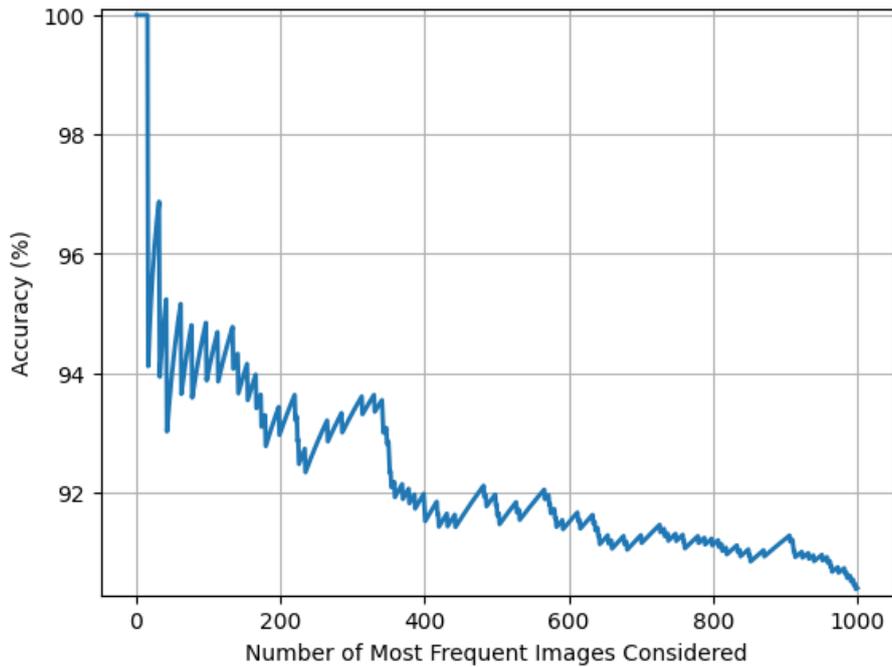


Figure 5.5: Percentage of positives within the n th most frequent image within the similars found by ULISSE over positive queries.

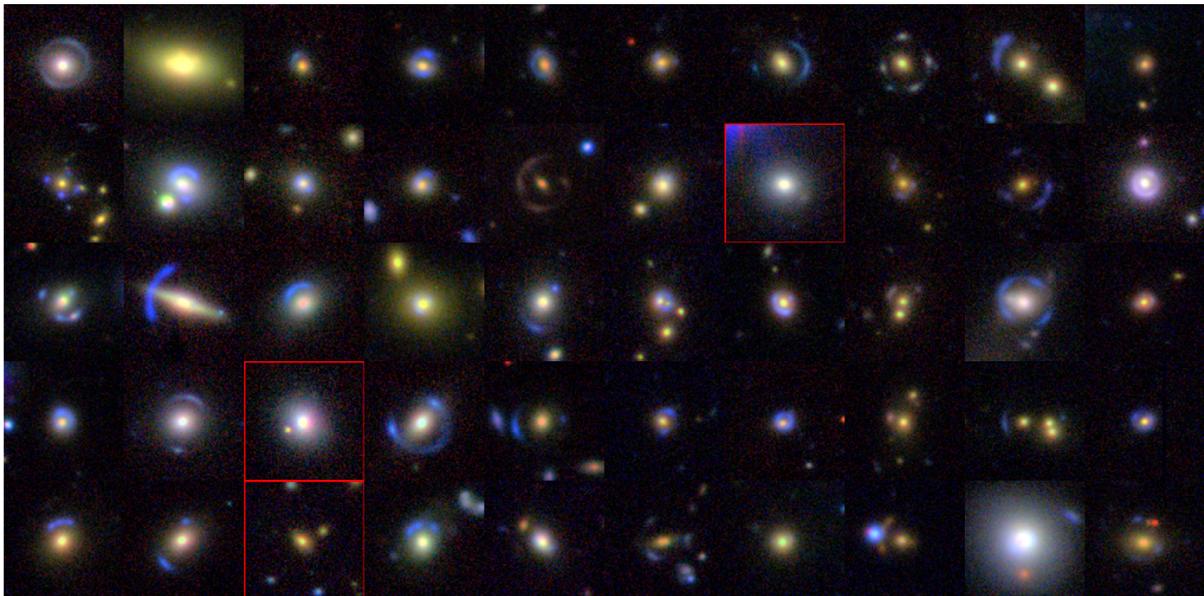


Figure 5.6: First fifty images for frequency within the list of similar, proposed by ULISSE, of the positive queries. The negatives have been framed within a red rectangle. These images have cutouts size of 101×101 pixels, different from the size of 40×40 pixels on which ULISSE has worked.

The accuracy of ULISSE within the most frequent images of the similars of the positive queries is displayed in Fig. 5.5. Within the first hundred most frequent candidates proposed by ULISSE on the positive queries, only six are negatives (94% accuracy); in the first two-hundred, 14 (93%).

5.2.1 Statistics of simulation parameters

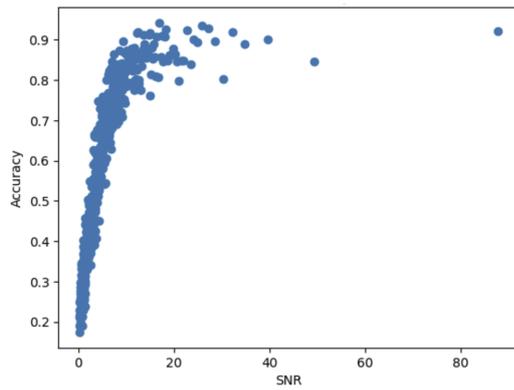
To understand how ULISSE works, we analyzed how the various simulation parameters affect ULISSE's performance, and the differences between the distribution of these parameters between the queries and the similars. The parameters used in the subsequent analysis are:

- **Signal-to-noise-ratio (SNR):** The local signal-to-noise ratio of the surface brightness peak of the arcs (3×3 pixels centered on the peak)
- **Alpha:** Ratio between the peak of the LRG flux and the arc flux.
- **Einstein Radius (R_{ein}):** Einstein radius used to model the mass distribution of the deflector (SIE).
- **Effective Radius (R_{eff}):** Effective radius of the intensity profile of the source (Sérsic).
- **Sérsic Index (n):** Sérsic index of the source.
- **mag_r , mag_g , mag_i :** Magnitudes of the source in different bands, as discussed in 4.2.2.

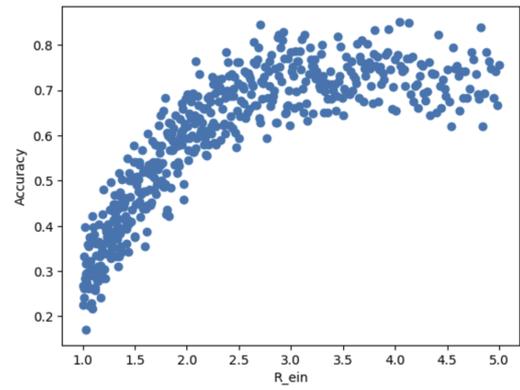
Each positive image is characterized by an array of parameters outlined above. Through ULISSE, a list of similar candidates is generated. Consequently, each parameter value of a query (e.g., the Einstein radius) is associated with an accuracy value. We plotted the values of a given parameter on the x -axis and the corresponding accuracies on the y -axis, averaging every 30 points. The results are presented in Figure 5.7.

Based on the plots, it is evident that the accuracy of ULISSE exhibits a strong correlation with both the Einstein Radius, the SNR, and alpha. These variables serve as crucial indicators of the gravitational arcs' dimensions and detectability, thus signifying that ULISSE finds the arc as the most important feature of a lensed object. On the other hand, parameters associated with the morphology of the source, such as the effective radius of the source and the Sérsic index, show no notable correlation with accuracy.

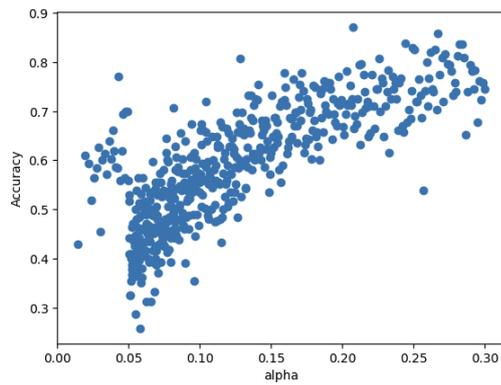
As expected, parameters connected to the magnitude of the source are strongly connected to the accuracy, as they represent the visibility of the source. The inverse relationship between the color ($g - r$) and ULISSE accuracy is interesting, notably as gravitational arcs tend to manifest a bluish color.



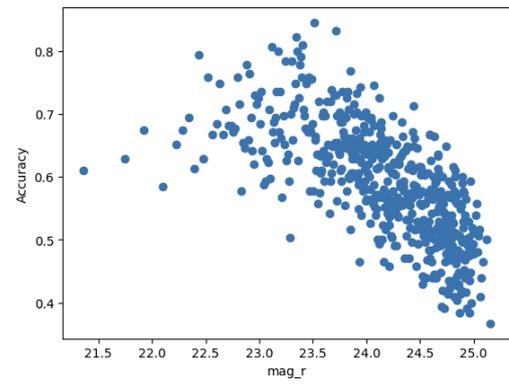
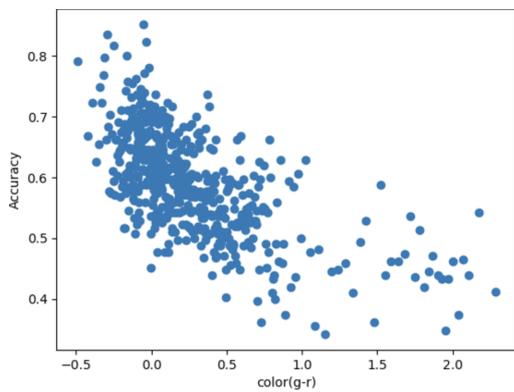
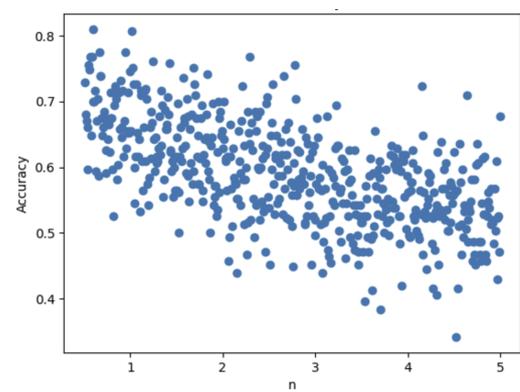
(a) SNR Parameters vs. Accuracy



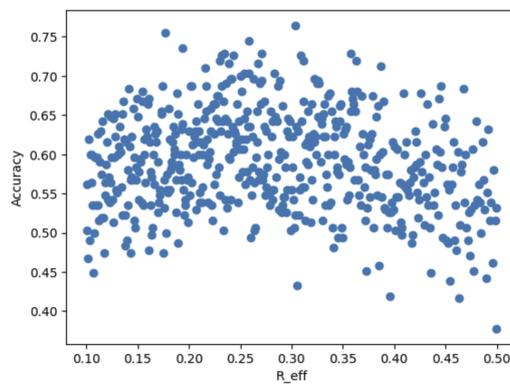
(b) Einstein Radius vs. Accuracy



(c) Parameter Alpha vs. Accuracy

(d) Magnitude in r band vs. Accuracy(e) Color ($g - r$) of the source vs. Accuracy

(f) Sérsic Index vs. Accuracy



(g) Effective Radius vs. Accuracy

Figure 5.7: Relationship between various simulation parameters and the accuracy of ULISSE. The plot shows the average accuracy over 30 data points.

5.3 Quasi-realistic case

The queries consist of the list of 1953 real galaxies with different probability to be a lens, identified in the KiDS (DR4) survey by [Petrillo et al. 2019b](#). Each query is associated with a score representing the likelihood of being a lens, determined through visual inspection by experts. The higher the score, the more the experts agree that the candidate is in fact a lens. Some examples are presented in Fig. 5.8

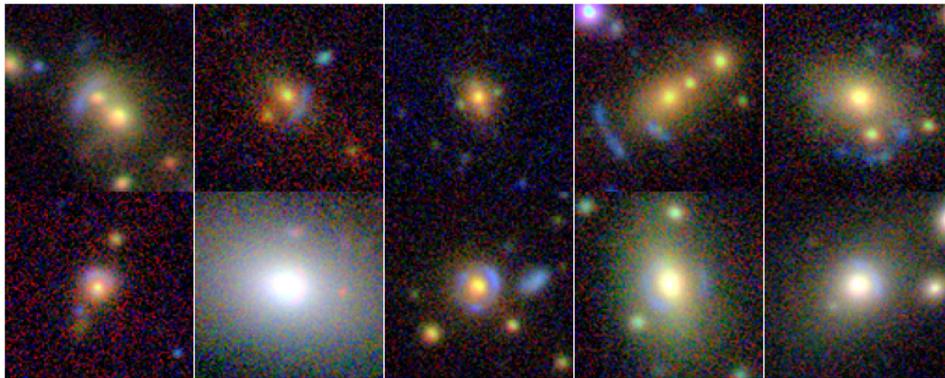


Figure 5.8: Ten examples of real lens candidates found by [Petrillo et al. 2019b](#)

A similarity search was conducted on the Dataset employing RGB and linear scale images. Moreover, we also consider cutouts size of 85 pixels. The case of a 40-pixel cutout was not explored, as, in this instance (in contrast to the case described in Sec 5.2), it resulted in the removal of essential features from the queries. The average accuracy of the queries, within the first 10 similars, is 35% with the cutout size of 85 pixels, whereas the accuracy increased to 52% with size of 101 pixels.

In Fig. 5.9 is presented the plot of the score against the average accuracy of ULISSE on queries with that specific score. In Table 5.1, the counts corresponding to each score are displayed. The

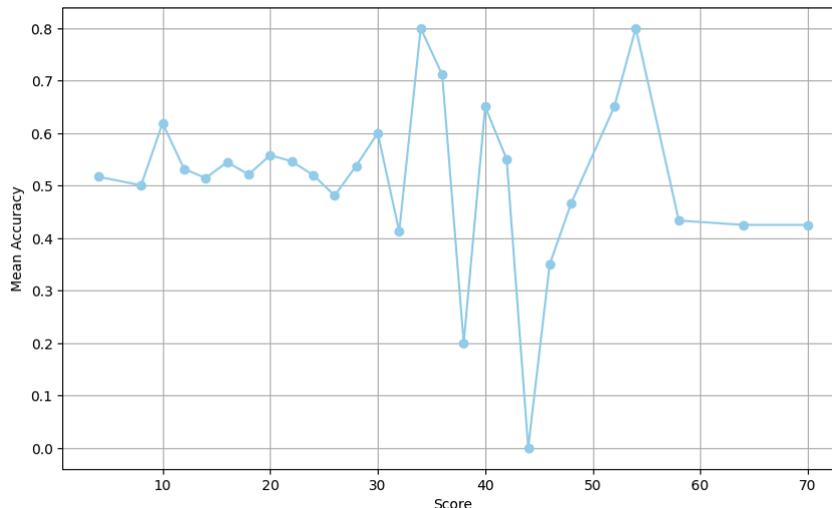
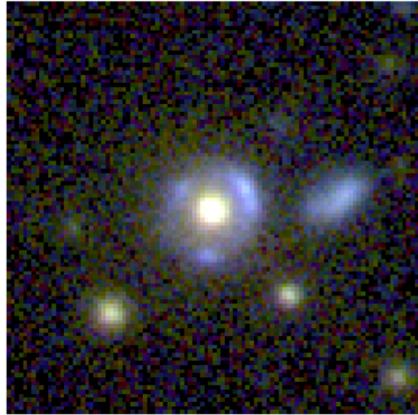


Figure 5.9: The score of the real lens candidates is on the x -axis, whereas the mean accuracy retrieved by ULISSE on the first ten candidates is on the y -axis.

variability observed in the data can be explained by analyzing the count distribution as depicted in Table 5.1. Notably, in the score interval of 38-54 the average number of images per score is 2.

Another test tried, was to change the number of nearest neighbors and here the best accuracy was found at k equal to 25, with the average accuracy of $\sim 53\%$. Furthermore, ULISSE

Query:



Loaded pretrained weights for efficientnet-b0
 The accuracy is: 100.0 %
 Random predictor accuracy: 29.04 %
 Nearest neighbors:

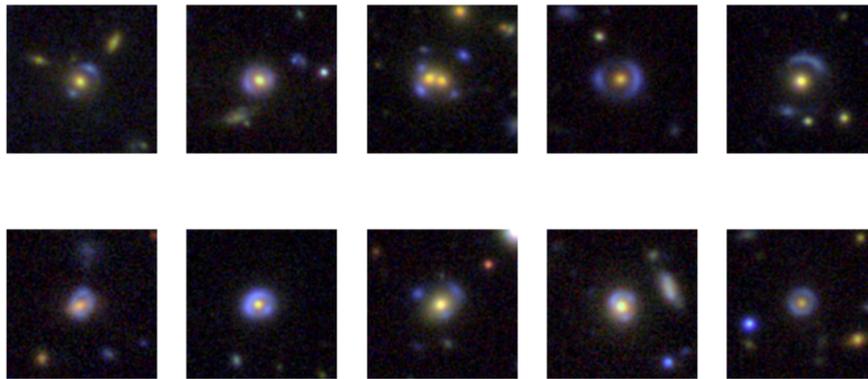


Figure 5.10: Example of how ULISSE works. The image at the top of the figure is one of the 1983 real lens candidates acting as query (with maximum score). The ten images at the bottom are the then most similar candidates found by ULISSE on the Dataset. The accuracy of ULISSE on this query is 100% meaning that all the similars are positives (simulated lenses). The random predictor accuracy represents the percentage of lenses in the whole Dataset

was tested in the reverse configuration. In this setup, the queries are simulated lenses, while the complementary dataset includes the negatives from the Dataset and the real lenses used as queries before. The average accuracy achieved was 1.8%, compared to the random classifier accuracy of 4.7%. In Fig. 5.11 is represented the plot of the average accuracy as function of k -nearest neighbors. The accuracy tends to slowly increasing toward the random classifier accuracy.

Table 5.1: Score counts

Score	4	8	10	12	14	16	18	20	22	24	26	28	30	32
Count	1061	409	16	161	28	81	38	38	26	15	21	16	16	8

Score	34	36	38	40	42	44	46	48	52	54	58	64	70
Count	4	9	2	2	6	1	2	3	2	1	9	4	4

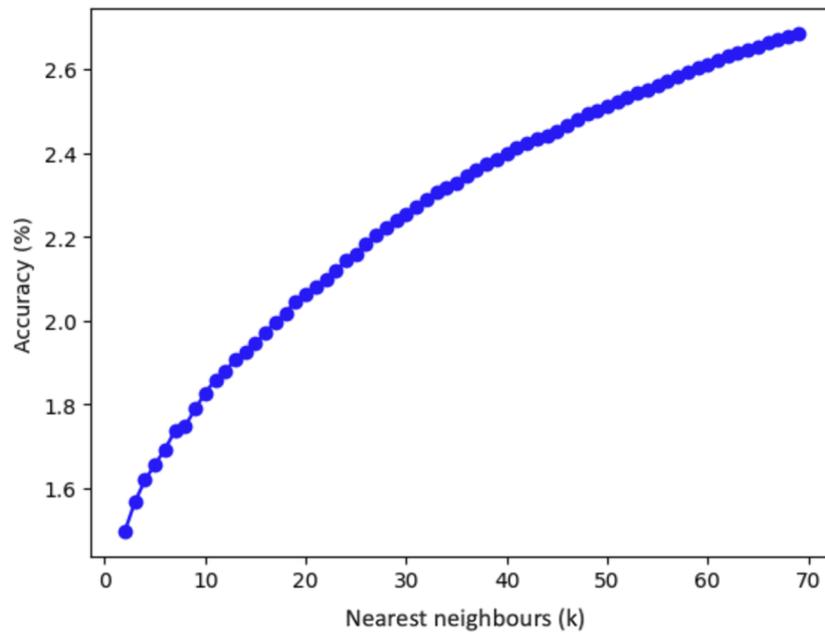


Figure 5.11: Accuracy vs. number of candidates. In this case, the queries are the simulated images (detailed in Sec 4.2.2) whereas the dataset is comprised by real lens candidates found by [Petrillo et al. \(2019b\)](#) and real negative images detailed in 4.2.1

Chapter 6

Discussion and Conclusions

The central scientific aim of this thesis focused on assessing the effectiveness of the ULISSE algorithm, a one-shot method that retrieves similar images to a given query, in the task of recognizing strong gravitational lenses. Gravitational lensing is a fascinating phenomenon predicted by Einstein’s General Relativity. When the light emitted by a background source travels through the gravitational field of a foreground object (the lens, e.g. a galaxy or a cluster of galaxies), is bent. The result of this phenomenon is the formation of amplified multiple-images or arcs/rings around the lens galaxy. Gravitational lensing is uniquely suited to provide valuable insights into the mass distribution within the (lens) galaxy, regardless of its luminosity. The magnifying power can also allow to observe faint sources, otherwise difficult to observe. Additionally, it serves as a vital method for probing the cosmological parameter H_0 , by measuring the times delays among multiple lensed-images, addressing the persistent challenge known as the “Hubble tension” (summarized in Verde et al. 2019) or potentially revealing new physics.

However, strong gravitational lenses are rare, leading to the development of numerous CNNs tailored for identifying these occurrences in extensive sky surveys (Jacobs et al. 2017; Petrillo et al. 2017; Pourrahmani et al. 2018; Petrillo et al. 2019a;b; He et al. 2020; Li et al. 2020; 2021; Gentile et al. 2023). Nevertheless, the use of simulated objects in training CNNs introduces inherent biases. Therefore, employing innovative methods such as ULISSE becomes useful.

ULISSE operates as a transfer learning approach, generating a list of similar images to a given query. To achieve this, ULISSE employs a large pre-trained CNN to decompose images into features, subsequently utilizing the k-NN algorithm for a similarity search between the feature vector of the query and the rest of the dataset. This process identifies the k -nearest neighbors in the feature vector space. ULISSE’s approach complements CNN methods for several reasons. While its accuracy is lower than that of CNNs, with CNNs reaching over 95% accuracy in the gravitational lensing task (Petrillo et al. 2017), ULISSE achieves approximately 60% accuracy (see Sec 5.2). Importantly, ULISSE operates without prior training, reducing resource and time requirements for simulations, and avoids bias, a significant advantage for recognizing outliers in the distribution of gravitational lenses. Furthermore, having the possibility to consider the similars of many lensed queries, then one can analyze the most frequent images within this subset and the accuracy can be incremented to 95% as seen in Sec 5.2. Results indicate that ULISSE benefits from the color content of images, with the best accuracy achieved using RGB images. Moreover, in terms of single-band images, the g -band, where the gravitational arc is more evident, achieves the highest accuracy. In the i -band, poor seeing conditions, coupled with reduced visibility of gravitational arcs compared to foreground lenses, result in the worst accuracy (see Fig. 5.2).

The choice of distance metric is not crucial, but the distance of a given image from its nearest neighbors in the feature space may be a relevant factor, considering the rarity of gravitational lenses. In the lens parameter analysis, high Einstein radius (indicating greater deflector mass) correlates with higher ULISSE accuracy, attributed to the larger and more easily recognizable

gravitational arcs associated with higher Einstein radii. This is also true for the SNR ratio of the surface brightness of the arc.

In the final section of our results, we observed that the accuracy of identified real lens candidates closely mirrors the outcomes observed with simulated queries. This suggests that our evaluation approach has been effective in preparing ULISSE for real case scenarios. Notably, when comparing the scores of individual real lens candidates with ULISSE’s mean accuracy on queries with that score, we noted an interesting pattern. ULISSE achieves its highest accuracy not with queries possessing the highest scores, but rather with those falling in the middle of the score range.

Moreover, we conducted a reverse scenario using simulated queries and the dataset, where ULISSE searches for candidates, comprising both real lens candidates and real negative images. This resulted in a notably lower average accuracy. One might think that this implies that ULISSE performs poorly when faced with significant imbalances in the dataset (only 4% positives). However, a more nuanced interpretation could be that, in situations with a low overall prevalence of positives, each lens may exhibit distinct characteristics, making it challenging for ULISSE to identify similarities. Therefore, our primary concern is the absolute number of lenses rather than their percentage within the dataset. This means that in real case scenarios it will not be a problem.

Given these findings, the ULISSE algorithm will be applied to the KiDS DR4 catalog of galaxies to identify new lens candidates and compare results with other CNNs pursuing similar objectives. ULISSE’s flexible architecture allows fast adaptation of the algorithm to new and more effective CNNs. The more features the CNNs extracts, the more ULISSE should be effective. On the other hand, as the dimension of the feature space grows, a change of metrics might be necessary.

Appendix A

Relativistic Deflection Angle

In General Relativity, the influence of a weak gravitational field ¹ on the surrounding space-time is described by the following metric tensor:

$$g_{\mu\nu} = \begin{bmatrix} 1 + 2\phi/c^2 & 0 & 0 & 0 \\ 0 & -1 - 2\phi/c^2 & 0 & 0 \\ 0 & 0 & -1 - 2\phi/c^2 & 0 \\ 0 & 0 & 0 & -1 - 2\phi/c^2 \end{bmatrix} \quad (\text{A.1})$$

A light ray traversing this gravitational field follows a null geodesic, characterized by the line element:

$$ds^2 = \left(1 + \frac{2\phi}{c^2}\right)(cdt)^2 - \left(1 - \frac{2\phi}{c^2}\right)dx^2 = 0. \quad (\text{A.2})$$

By a straightforward algebraic manipulation, the speed of light propagation through the gravitational field can be determined as:

$$v = \frac{dx}{dt} = c \sqrt{\frac{1 + \frac{2\phi}{c^2}}{1 - \frac{2\phi}{c^2}}} \approx c \left(1 + \frac{2\phi}{c^2}\right). \quad (\text{A.3})$$

Recalling the definition of the refractive index n , we can express it as:

$$n = \frac{c}{v} = \frac{1}{\sqrt{1 + \phi/c^2}} = 1 - \frac{2\phi}{c^2}. \quad (\text{A.4})$$

Using this relation, we can infer that the gravitational field acts as an optical medium characterized by a refractive index n contingent on the potential.

Furthermore, it is important to highlight that, given the non-uniform nature of the gravitational field, the corresponding refractive index becomes a function of the position $n(x)$. Following *Fermat's Principle*, the path taken by a light ray traveling from point A to point B through an optical medium can be determined by solving the variational equation:

$$\delta \int_A^B n[x(\ell)] d\ell = 0. \quad (\text{A.5})$$

This can be accomplished by employing the Euler-Lagrange equations. Initially, a reparametrization of the trajectory is required using a new curve parameter λ :

$$d\ell = \frac{dx}{d\lambda} d\lambda. \quad (\text{A.6})$$

¹A gravitational field is considered weak if it does not induce relativistic effects on the subjected matter. This property is defined by the relation: $\phi/c^2 \ll 1$.

Consequently, Equation A.5 can be reformulated as:

$$\delta \int_{\lambda_A}^{\lambda_B} n[x(\lambda)] \frac{dx}{d\lambda} d\lambda. \quad (\text{A.7})$$

Next, defining a Lagrangian function L as:

$$L(x, \dot{x}, \lambda) = n[x(\lambda)] \frac{dx}{d\lambda}, \quad (\text{A.8})$$

with $\dot{x} = \frac{dx}{d\lambda}$. The Euler-Lagrange equations can be expressed as:

$$\frac{d}{d\lambda} \left(\frac{\partial L}{\partial \dot{x}} \right) - \frac{\partial L}{\partial x} = 0. \quad (\text{A.9})$$

From Equation A.8, it follows:

$$\frac{\partial L}{\partial x} = |\dot{x}| \frac{\partial n}{\partial x} = (\nabla n) |\dot{x}| \quad \frac{\partial L}{\partial \dot{x}} = n \dot{x}. \quad (\text{A.10})$$

We can appropriately choose the λ parameter to have a normalized tangent vector \dot{x} . By denoting $e = \dot{x}$ this versor, the Euler-Lagrange equations can be rewritten as:

$$\frac{d}{d\lambda} (ne) - \nabla n = 0, \quad (\text{A.11})$$

with some algebra we can obtain:

$$n \dot{e} = \nabla n - e(\nabla n \cdot e), \quad (\text{A.12})$$

in the second term of this expression, the derivative of the refraction index along the light path can be recognized. Thus, the entire equation can be rewritten as:

$$\dot{e} = \frac{1}{n} \nabla_{\perp} n = \nabla \ln n. \quad (\text{A.13})$$

Finally, using Equation A.4 and employing the weak-field approximation, we can obtain:

$$\dot{e} \approx -\frac{2}{c^2} \nabla_{\perp} \phi. \quad (\text{A.14})$$

Thus, the total deflection angle can be evaluated as:

$$\alpha = \frac{2}{c^2} \int_{\lambda_A}^{\lambda_B} \nabla_{\perp} \phi \cdot d\lambda. \quad (\text{A.15})$$

Equation A.15 cannot be evaluated analytically since the integral should be performed along the true (unknown) light path. Therefore, the Born approximation is commonly employed: in the weak-field approximation, it is reasonable to assume a small deflection angle for the light trajectory. Thus, we can approximate the actual path followed by the light ray with a straight line. With this approximation, Equation A.15 can be rewritten as:

$$\alpha = \frac{2}{c^2} \int_{-1}^{+1} \nabla_{\perp}(\phi) \cdot dz. \quad (\text{A.16})$$

Starting from this expression, it is simple to derive the deflection angle for a point-like mass M :

$$\alpha = \frac{4GM}{bc^2}, \quad (\text{A.17})$$

which, as described in Section 2.1, differs from the classical estimation exactly by a factor of 2.

Appendix B

Sérsic profile

The $R^{1/4}$ law, introduced by [de Vaucouleurs \(1948\)](#), was the fitting function used for decades to model the surface brightness $I(R)$ of galaxies. This profile has no free parameters and makes use of two physical scales: a characteristic linear scale (R_e) and a luminosity factor (I_e). It was demonstrated early that the surface brightness of elliptical galaxies is well-fitted by this law (e.g., [Kormendy 1977](#); [de Vaucouleurs and Capaccioli 1979](#)). However, today's model of choice for describing galaxies is the Sérsic profile ([Sérsic 1963](#); [Sersic 1968](#)), which is a generalization of the $R^{1/4}$ law. The key advantages of the model are two:

- Describe the observed stellar distributions that have a range of central concentrations,
- Fit well the data for the entire radial extent.

The success comes from the fact that the Sérsic model requires additional fitting parameters which provide an extra degree of freedom (see more at [Graham 2011](#)). It can be expressed as an intensity profile, such as:

$$I(R) = I_e \exp -b_n \left[\left(\frac{R}{R_e} \right)^{1/n} - 1 \right], \quad (\text{B.1})$$

where I_e is the intensity at the effective radius R_e that encloses half of the total light from the model. The constant b_n is defined in terms of the parameter n which describes the “shape” of the source. The model reproduces the exponential model when $n = 1$ and the de Vaucouleurs' $R^{1/4}$ law for $n = 4$.

The luminosity L , at radius R , can be obtained by integrating Eq. [B.1](#) over an area $A = \pi R^2$. Substituting R with $x = b_n \left(\frac{R}{R_e} \right)^{1/n}$, it yields:

$$L(R) = I_e R_e^2 2\pi n \frac{e^{b_n}}{(b_n)^{2n}} \gamma(2n, x) \quad (\text{B.2})$$

where,

$$\gamma(2n, x) = \int_0^x e^{-t} t^{2n-1} \quad (\text{B.3})$$

Replacing $\gamma(2n, x)$ with $\Gamma(2n)$ in Eq. [\(B.2\)](#) gives the value of L_{tot} ([Ciotti 1991](#)). Therefore, the value of the constant b_n is such that:

$$\Gamma(2n) = 2\gamma(2n, b_n), \quad (\text{B.4})$$

where Γ is the complete gamma function ([Ciotti 1991](#)). The exact value of b_n can only be obtained numerically.

Multiplying the negative logarithm of the luminosity profile, Eq. [\(B.2\)](#), by 2.5 gives the enclosed magnitude profile, known as the “curve of growth”,

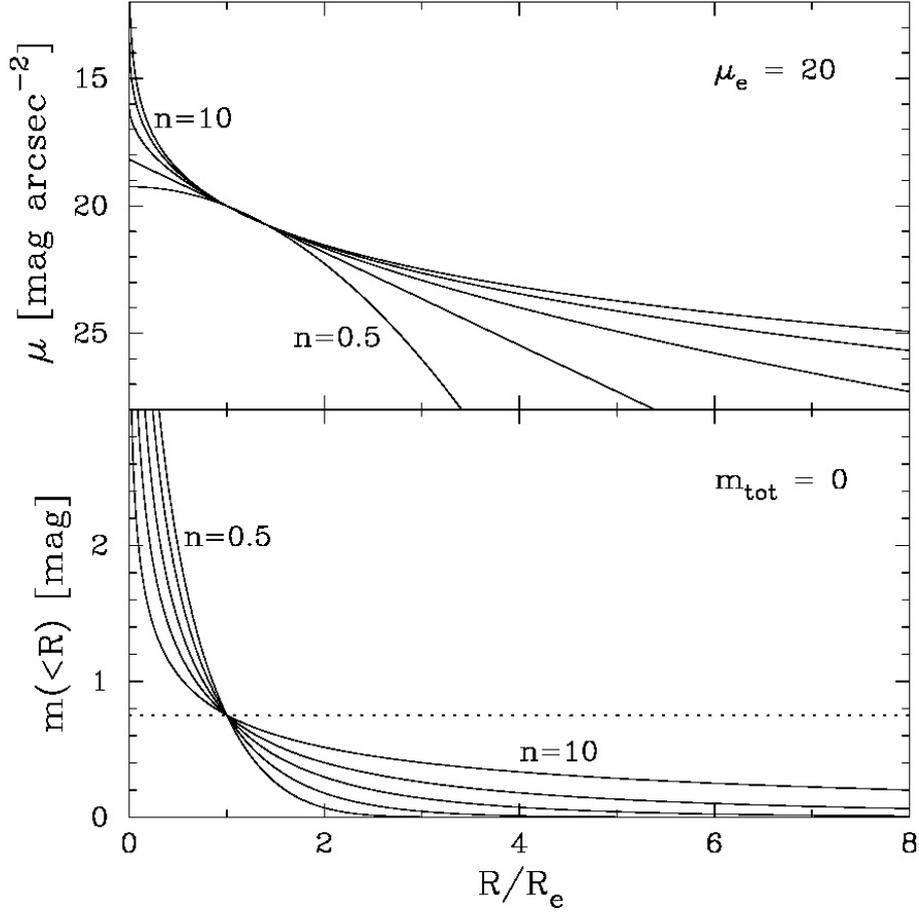


Figure B.1: Top panel: Sérsic surface brightness profiles, Eq. B.6 for $n = 0.5, 1, 2, 4, 10$, the profiles are normalized at $\mu_e = 20 \text{ mag arcsec}^{-2}$

Bottom panel: Sérsic aperture magnitude profiles, Eq. B.5 normalized such that the total magnitude equals zero. The dotted line is shifted by 0.75 mag from the total magnitude. Original source: [Graham and Driver 2005](#)

$$m(R) = \mu_e - 5 \log R_e - 2.5 \log \left[2\pi n \frac{e^{b_n}}{(b_n)^{2n} \gamma(2n, x)} \right]. \quad (\text{B.5})$$

Here, μ indicate the surface brightness in mag arcsec^{-2} . As R tends to infinity the equation yields the total apparent magnitude and, consequently, $\gamma(2n, x) \rightarrow \Gamma(2n)$, see Fig. B.1.

Multiplying the negative logarithm of Eq. B.1 by 2.5 yields the surface brightness profile,

$$\mu(R) = \mu_e + \frac{2.5b_n}{\ln(10)} \left[\left(\frac{R}{R_e} \right)^{1/n} - 1 \right]. \quad (\text{B.6})$$

Appendix C

Point Spread Function

The degradation of surface brightness measurements of galaxies, resulting from atmospheric blurring and imperfections in telescope optics (known as seeing), is a well-recognized challenge. Characterizing the seeing phenomenon is accomplished through the point spread function (PSF).

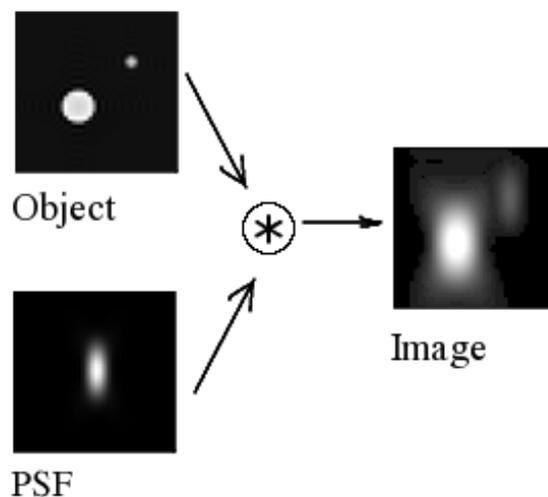


Figure C.1: The image seen by the imaging device is obtained through the convolution of the surface brightness of the astronomical object with the PSF of the telescope.

The PSF describes the response of an imaging system to a point source. Therefore, empirically, PSF can be determined by studying the scattering of stellar light (Moffat 1969; King 1971).

In general, Gaussian functions, linear superpositions of Gaussian functions, or Moffat functions effectively describe PSFs.

The Moffat PSF presents two notable advantages over the Gaussian PSF (see more at Trujillo et al. 2001):

- It allows effective fitting of “wings” observed in stellar profiles.
- It is numerically well-behaved when dealing with narrow PSFs.

For these reasons, the Moffat function has been widely used to model the PSF. This profile is mathematically described by:

$$\text{PSF}(r) = \frac{\beta - 1}{\pi\alpha^2} \left[1 + \left(\frac{r}{\alpha} \right)^2 \right]^{-\beta}. \quad (\text{C.1})$$

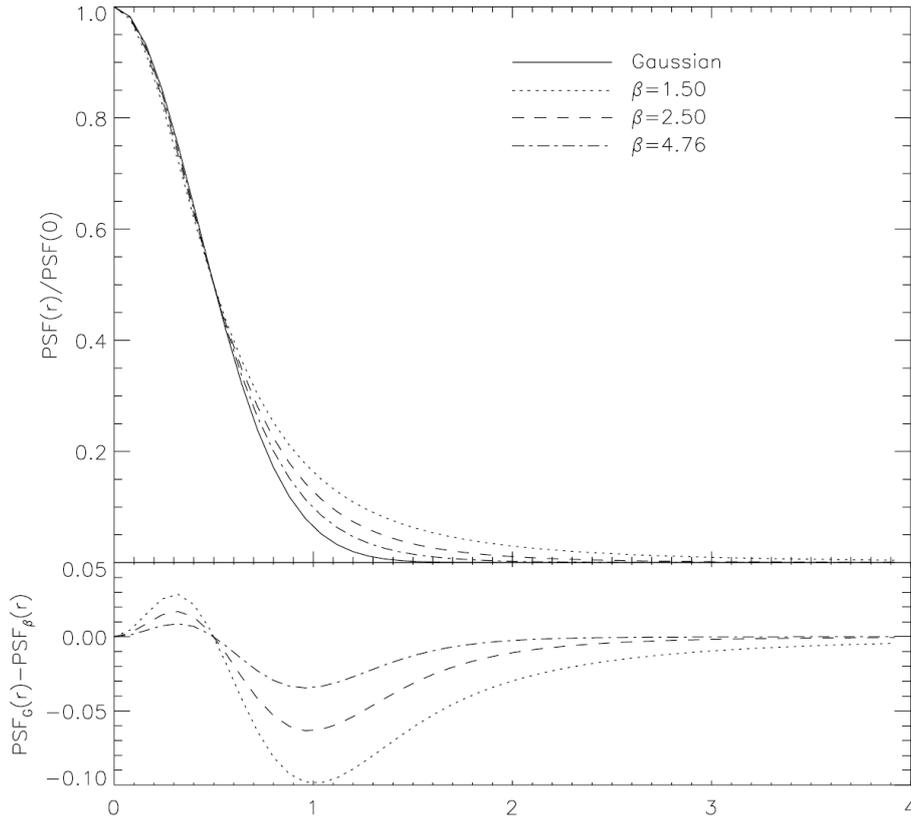


Figure C.2: The top panel illustrates the normalized intensity profiles of Gaussian and Moffat PSF functions, showcasing their variations with different β values. The profiles are plotted against the PSF radius in FWHM units. The bottom panel displays the discrepancy between the normalized Moffat PSF ($PSF_\beta(r)$) and the normalized Gaussian PSF ($PSF_G(r)$). Original source: [Trujillo et al. 2001](#)

Here, the full width at maximum height (FWMH) is given by $FWMH = 2\alpha\sqrt{2^{1/\beta} - 1}$, where $PSF(FWHM/2) = \frac{1}{2}PSF(0)$ and the total flux is normalized to one. Fig. C.2 illustrates that as β increases, the Moffat function approximates the core of the Gaussian profile. In fact, for $\beta \rightarrow \infty$, the Moffat function reduces to a Gaussian profile.

The relevant features of the PSF, in KSB models ([Kaiser et al. 1995](#)) and its variants ([Luppino and Kaiser 1997](#); [Hoekstra et al. 1998](#)), are its ellipticity and size, which are estimated from the second-order geometrical moments of the PSF image. The weighted quadruple moment reads as:

$$Q_{ij} = \int d^2\theta W(\theta)\theta_i\theta_j S(\theta), \quad (C.2)$$

where angles are measured relative to the object position as determined in the detection phase. $W(\theta)$ represents a Gaussian weight function of the scale length, which is some measures of the PSF size, such as the half-light radius. Then, PSF ellipticity ϵ has two components:

$$\epsilon_1 = \frac{Q_{11} - Q_{22}}{Q_{11} + Q_{22}}, \quad \epsilon_2 = \frac{2Q_{11}}{Q_{11} + Q_{22}}. \quad (C.3)$$

As usual in large field detectors, the PSF is a function of the position across the field-of-view. For example, the ellipticity pattern across the field-of-view in KiDS is represented in Fig. C.3

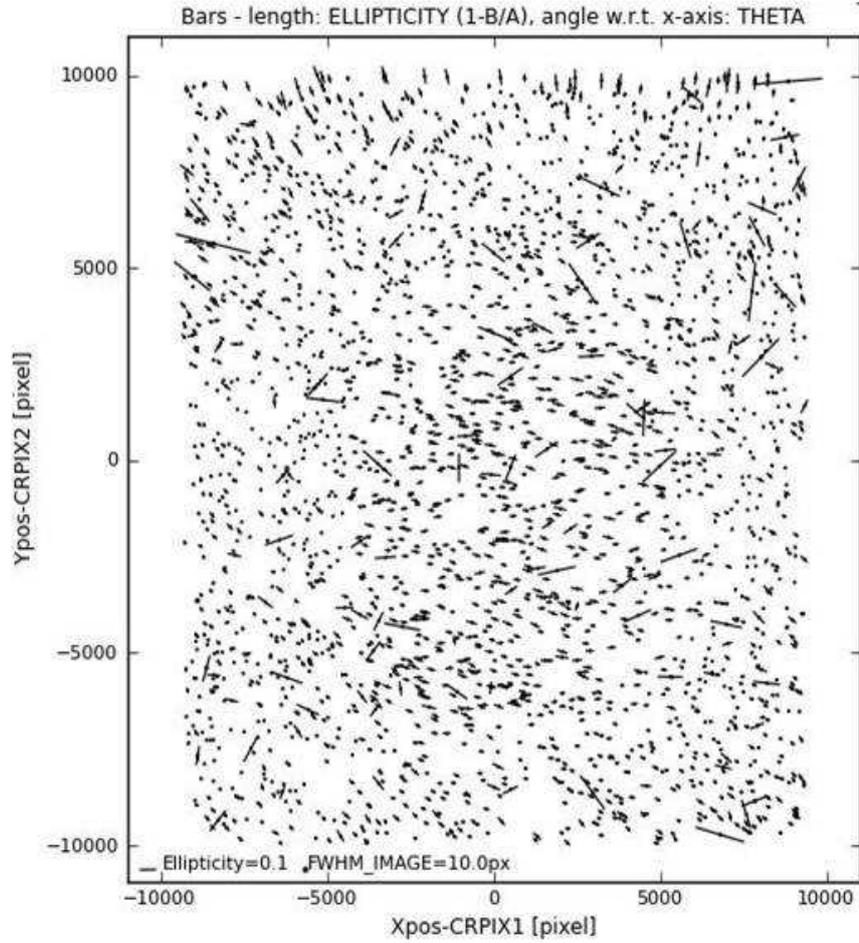


Figure C.3: PSF anisotropy within the coadded KIDS 129.0 -0.5 in r-band. The elongation is aligned in a specific direction on the borders but random in the middle of the image. Original source: [Kuijken et al. 2019](#)

Bibliography

- Amaro, V., Cavuoti, S., Brescia, M., et al., 2019. Statistical analysis of probability density functions for photometric redshifts through the KiDS-ESO-DR3 galaxies. *MNRAS* 482, 3116–3134. doi:[10.1093/mnras/sty2922](https://doi.org/10.1093/mnras/sty2922), [arXiv:1810.09777](https://arxiv.org/abs/1810.09777).
- Asgari, M., Heymans, C., Hildebrandt, H., et al., 2019. Consistent cosmic shear in the face of systematics: a B-mode analysis of KiDS-450, DES-SV and CFHTLenS. *A&A* 624, A134. doi:[10.1051/0004-6361/201834379](https://doi.org/10.1051/0004-6361/201834379), [arXiv:1810.02353](https://arxiv.org/abs/1810.02353).
- Auger, M.W., Treu, T., Bolton, A.S., et al., 2010. The Sloan Lens ACS Survey. X. Stellar, Dynamical, and Total Mass Correlations of Massive Early-type Galaxies. *Astrophysical Journal* 724, 511–525. doi:[10.1088/0004-637X/724/1/511](https://doi.org/10.1088/0004-637X/724/1/511), [arXiv:1007.2880](https://arxiv.org/abs/1007.2880).
- Baron, D., Poznanski, D., 2017. The weirdest SDSS galaxies: results from an outlier detection algorithm. *MNRAS* 465, 4530–4555. doi:[10.1093/mnras/stw3021](https://doi.org/10.1093/mnras/stw3021), [arXiv:1611.07526](https://arxiv.org/abs/1611.07526).
- Bellman, R., 1957. *Dynamic programming*. Princeton university press .
- Benítez, N., 2000. Bayesian Photometric Redshift Estimation. *Astrophysical Journal* 536, 571–583. doi:[10.1086/308947](https://doi.org/10.1086/308947), [arXiv:astro-ph/9811189](https://arxiv.org/abs/astro-ph/9811189).
- Bentley, J.L., 1975. Multidimensional binary search trees used for associative searching. *Commun. ACM* 18, 509–517. URL: <https://api.semanticscholar.org/CorpusID:13091446>.
- Bertin, E., Arnouts, S., 1996. SExtractor: Software for source extraction. *A&AS* 117, 393–404. doi:[10.1051/aas:1996164](https://doi.org/10.1051/aas:1996164).
- Binney, J., Merrifield, M., Wegner, G.A., 2000. *Galactic Astronomy*. doi:[10.1119/1.19402](https://doi.org/10.1119/1.19402).
- Birrer, S., Millon, M., Sluse, D., et al., 2022. Time-Delay Cosmography: Measuring the Hubble Constant and other cosmological parameters with strong gravitational lensing. *arXiv e-prints* doi:[10.48550/arXiv.2210.10833](https://doi.org/10.48550/arXiv.2210.10833), [arXiv:2210.10833](https://arxiv.org/abs/2210.10833).
- Bishop, C.M., 2006. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg.
- Blandford, R.D., Narayan, R., 1992. Cosmological applications of gravitational lensing. *ARA&A* 30, 311–358. doi:[10.1146/annurev.astro.30.1.311](https://doi.org/10.1146/annurev.astro.30.1.311).
- Bolton, A.S., Burles, S., Koopmans, L.V.E., et al., 2006. The Sloan Lens ACS Survey. I. A Large Spectroscopically Selected Sample of Massive Early-Type Lens Galaxies. *Astrophysical Journal* 638, 703–724. doi:[10.1086/498884](https://doi.org/10.1086/498884), [arXiv:astro-ph/0511453](https://arxiv.org/abs/astro-ph/0511453).
- Boroson, T.A., Green, R.F., 1992. The Emission-Line Properties of Low-Redshift Quasi-stellar Objects. *ApJS* 80, 109. doi:[10.1086/191661](https://doi.org/10.1086/191661).
- Brescia, M., Cavuoti, S., D’Abrusco, R., et al., 2013. Photometric Redshifts for Quasars in Multi-band Surveys. *Astrophysical Journal* 772, 140. doi:[10.1088/0004-637X/772/2/140](https://doi.org/10.1088/0004-637X/772/2/140), [arXiv:1305.5641](https://arxiv.org/abs/1305.5641).

- Brescia, M., Djorgovski, S., Feigelson, E.D., et al., 2017. *Astroinformatics*. volume 325.
- Brouwer, M.M., Cacciato, M., Dvornik, A., et al., 2016. Dependence of GAMA galaxy halo masses on the cosmic web environment from 100 deg² of KiDS weak lensing data. *MNRAS* 462, 4451–4463. doi:[10.1093/mnras/stw1602](https://doi.org/10.1093/mnras/stw1602), [arXiv:1604.07233](https://arxiv.org/abs/1604.07233).
- Brown, T.B., Mann, B., Ryder, N., et al., 2020. Language Models are Few-Shot Learners. doi:[10.48550/arXiv.2005.14165](https://doi.org/10.48550/arXiv.2005.14165), [arXiv:2005.14165](https://arxiv.org/abs/2005.14165).
- Cai, L., Gao, J., Zhao, D., 2020. A review of the application of deep learning in medical image classification and segmentation. *Annals of translational medicine* 8.
- Capozziello, S., de Laurentis, M., 2011. Extended Theories of Gravity. *Phys. Rep.* 509, 167–321. doi:[10.1016/j.physrep.2011.09.003](https://doi.org/10.1016/j.physrep.2011.09.003), [arXiv:1108.6266](https://arxiv.org/abs/1108.6266).
- Castro-Ginard, A., Jordi, C., Luri, X., et al., 2018. A new method for unveiling open clusters in Gaia. New nearby open clusters confirmed by DR2. *A&A* 618, A59. doi:[10.1051/0004-6361/201833390](https://doi.org/10.1051/0004-6361/201833390), [arXiv:1805.03045](https://arxiv.org/abs/1805.03045).
- Cavuoti, S., Tortora, C., Brescia, M., et al., 2017. A cooperative approach among methods for photometric redshifts estimation: an application to KiDS data. *MNRAS* 466, 2039–2053. doi:[10.1093/mnras/stw3208](https://doi.org/10.1093/mnras/stw3208), [arXiv:1612.02173](https://arxiv.org/abs/1612.02173).
- Chae, K.H., 2003. The Cosmic Lens All-Sky Survey: statistical strong lensing, cosmological parameters, and global properties of galaxy populations. *MNRAS* 346, 746–772. doi:[10.1111/j.1365-2966.2003.07092.x](https://doi.org/10.1111/j.1365-2966.2003.07092.x), [arXiv:astro-ph/0211244](https://arxiv.org/abs/astro-ph/0211244).
- Chatterjee, S., Koopmans, L.V.E., 2018. The inner mass power spectrum of galaxies using strong gravitational lensing: beyond linear approximation. *MNRAS* 474, 1762–1772. doi:[10.1093/mnras/stx2674](https://doi.org/10.1093/mnras/stx2674), [arXiv:1710.03075](https://arxiv.org/abs/1710.03075).
- Chevance, M., Weijmans, A.M., Damjanov, I., et al., 2012. On the Shapes and Structures of High-redshift Compact Galaxies. *ApJ* 754, L24. doi:[10.1088/2041-8205/754/2/L24](https://doi.org/10.1088/2041-8205/754/2/L24), [arXiv:1206.3579](https://arxiv.org/abs/1206.3579).
- Ciotti, L., 1991. Stellar systems following the R1/m luminosity law. *A&A* 249, 99–106.
- Collett, T.E., 2015. The Population of Galaxy-Galaxy Strong Lenses in Forthcoming Optical Imaging Surveys. *Astrophysical Journal* 811, 20. doi:[10.1088/0004-637X/811/1/20](https://doi.org/10.1088/0004-637X/811/1/20), [arXiv:1507.02657](https://arxiv.org/abs/1507.02657).
- Collister, A.A., Lahav, O., 2004. ANNz: Estimating Photometric Redshifts Using Artificial Neural Networks. *PASP* 116, 345–351. doi:[10.1086/383254](https://doi.org/10.1086/383254), [arXiv:astro-ph/0311058](https://arxiv.org/abs/astro-ph/0311058).
- Congdon, A.B., Keeton, C.R., 2018. Principles of Gravitational Lensing: Light Deflection as a Probe of Astrophysics and Cosmology. doi:[10.1007/978-3-030-02122-1](https://doi.org/10.1007/978-3-030-02122-1).
- Connolly, A.J., Peterson, J., Jernigan, J.G., et al., 2010. Simulating the LSST system, in: Angeli, G.Z., Dierickx, P. (Eds.), *Modeling, Systems Engineering, and Project Management for Astronomy IV*, p. 77381O. doi:[10.1117/12.857819](https://doi.org/10.1117/12.857819).
- Connolly, A.J., Szalay, A.S., Bershad, M.A., et al., 1995. Spectral Classification of Galaxies: an Orthogonal Approach. *AJ* 110, 1071. doi:[10.1086/117587](https://doi.org/10.1086/117587), [arXiv:astro-ph/9411044](https://arxiv.org/abs/astro-ph/9411044).
- Cover, T.M., Hart, P.E., 1967. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* 13, 21–27. URL: <https://api.semanticscholar.org/CorpusID:5246200>.

- de Jong, J.T.A., Kuijken, K., Applegate, D., et al., 2013. The Kilo-Degree Survey. *The Messenger* 154, 44–46.
- De Lucia, G., Springel, V., White, S.D.M., et al., 2006. The formation history of elliptical galaxies. *MNRAS* 366, 499–509. doi:[10.1111/j.1365-2966.2005.09879.x](https://doi.org/10.1111/j.1365-2966.2005.09879.x), [arXiv:astro-ph/0509725](https://arxiv.org/abs/astro-ph/0509725).
- de Vaucouleurs, G., 1948. Recherches sur les Nebuleuses Extragalactiques. *Annales d’Astrophysique* 11, 247.
- de Vaucouleurs, G., Capaccioli, M., 1979. Luminosity distribution in galaxies. I. The elliptical galaxy NGC 3379 as a luminosity distribution standard. *ApJS* 40, 699–731. doi:[10.1086/190602](https://doi.org/10.1086/190602).
- Deane, R.P., Rawlings, S., Garrett, M.A., et al., 2013. The preferentially magnified active nucleus in IRAS F10214+4724 - III. VLBI observations of the radio core. *MNRAS* 434, 3322–3336. doi:[10.1093/mnras/stt1241](https://doi.org/10.1093/mnras/stt1241), [arXiv:1307.6566](https://arxiv.org/abs/1307.6566).
- Delli Veneri, M., Cavuoti, S., Brescia, M., et al., 2019. Star formation rates for photometric samples of galaxies using machine learning methods. *MNRAS* 486, 1377–1391. doi:[10.1093/mnras/stz856](https://doi.org/10.1093/mnras/stz856), [arXiv:1902.02522](https://arxiv.org/abs/1902.02522).
- Deng, J., Dong, W., Socher, R., et al., 2009. Imagenet: A large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, pp. 248–255. doi:[10.1109/CVPR.2009.5206848](https://doi.org/10.1109/CVPR.2009.5206848).
- Dewdney, P.E., Hall, P.J., Schilizzi, R.T., Lazio, T.J.L.W., 2009. The Square Kilometre Array. *IEEE Proceedings* 97, 1482–1496. doi:[10.1109/JPROC.2009.2021005](https://doi.org/10.1109/JPROC.2009.2021005).
- Djorgovski, S.G., Graham, M.J., Donalek, C., et al., 2016. Real-Time Data Mining of Massive Data Streams from Synoptic Sky Surveys. *arXiv e-prints*, [arXiv:1601.04385](https://arxiv.org/abs/1601.04385)doi:[10.48550/arXiv.1601.04385](https://doi.org/10.48550/arXiv.1601.04385), [arXiv:1601.04385](https://arxiv.org/abs/1601.04385).
- Doorenbos, L., Torbaniuk, O., Cavuoti, S., et al., 2022. ULISSE: A tool for one-shot sky exploration and its application for detection of active galactic nuclei. *A&A* 666, A171. doi:[10.1051/0004-6361/202243900](https://doi.org/10.1051/0004-6361/202243900), [arXiv:2208.10984](https://arxiv.org/abs/2208.10984).
- Dvornik, A., Cacciato, M., Kuijken, K., et al., 2017. A KiDS weak lensing analysis of assembly bias in GAMA galaxy groups. *MNRAS* 468, 3251–3265. doi:[10.1093/mnras/stx705](https://doi.org/10.1093/mnras/stx705), [arXiv:1703.06657](https://arxiv.org/abs/1703.06657).
- Dyson, F.W., Eddington, A.S., Davidson, C., 1920. A Determination of the Deflection of Light by the Sun’s Gravitational Field, from Observations Made at the Total Eclipse of May 29, 1919. *Philosophical Transactions of the Royal Society of London Series A* 220, 291–333. doi:[10.1098/rsta.1920.0009](https://doi.org/10.1098/rsta.1920.0009).
- Edge, A., Sutherland, W., Kuijken, K., et al., 2013. The VISTA Kilo-degree Infrared Galaxy (VIKING) Survey: Bridging the Gap between Low and High Redshift. *The Messenger* 154, 32–34.
- Einstein, A., 1916. Die Grundlage der allgemeinen Relativitätstheorie. *Annalen der Physik* 354, 769–822. doi:[10.1002/andp.19163540702](https://doi.org/10.1002/andp.19163540702).
- Einstein, A., 1936. Lens-Like Action of a Star by the Deviation of Light in the Gravitational Field. *Science* 84, 506–507. doi:[10.1126/science.84.2188.506](https://doi.org/10.1126/science.84.2188.506).

- Eisenstein, D.J., Annis, J., Gunn, J.E., et al., 2001. Spectroscopic Target Selection for the Sloan Digital Sky Survey: The Luminous Red Galaxy Sample. *AJ* 122, 2267–2280. doi:[10.1086/323717](https://doi.org/10.1086/323717), [arXiv:astro-ph/0108153](https://arxiv.org/abs/astro-ph/0108153).
- Fix, E., Hodges, J.L., 1951. Discriminatory analysis - nonparametric discrimination: Consistency properties. *International Statistical Review* 57, 238. URL: <https://api.semanticscholar.org/CorpusID:120323383>.
- Foëx, G., Motta, V., Limousin, M., et al., 2013. SARCS strong-lensing galaxy groups. I. Optical, weak lensing, and scaling laws. *A&A* 559, A105. doi:[10.1051/0004-6361/201321112](https://doi.org/10.1051/0004-6361/201321112), [arXiv:1308.4674](https://arxiv.org/abs/1308.4674).
- Fukugita, M., Futamase, T., Kasai, M., Turner, E.L., 1992. Statistical Properties of Gravitational Lenses with a Nonzero Cosmological Constant. *Astrophysical Journal* 393, 3. doi:[10.1086/171481](https://doi.org/10.1086/171481).
- Gentile, F., Tortora, C., Covone, G., et al., 2023. LEMON: LENS MOdelling with Neural networks - I. Automated modelling of strong gravitational lenses with Bayesian Neural Networks. *MNRAS* 522, 5442–5455. doi:[10.1093/mnras/stad1325](https://doi.org/10.1093/mnras/stad1325), [arXiv:2210.10793](https://arxiv.org/abs/2210.10793).
- Gong, Y., Liu, X., Cao, Y., et al., 2019. Cosmology from the Chinese Space Station Optical Survey (CSS-OS). *Astrophysical Journal* 883, 203. doi:[10.3847/1538-4357/ab391e](https://doi.org/10.3847/1538-4357/ab391e), [arXiv:1901.04634](https://arxiv.org/abs/1901.04634).
- Graham, A.W., 2011. A review of elliptical and disc galaxy structure, and modern scaling laws. *arXiv: Cosmology and Nongalactic Astrophysics*, 91–140 URL: <https://api.semanticscholar.org/CorpusID:117887959>.
- Graham, A.W., Driver, S.P., 2005. A Concise Reference to (Projected) Sérsic $R^{1/n}$ Quantities, Including Concentration, Profile Slopes, Petrosian Indices, and Kron Magnitudes. *PASA* 22, 118–127. doi:[10.1071/AS05001](https://doi.org/10.1071/AS05001), [arXiv:astro-ph/0503176](https://arxiv.org/abs/astro-ph/0503176).
- He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. *arXiv e-prints*, [arXiv:1502.01852](https://arxiv.org/abs/1502.01852)doi:[10.48550/arXiv.1502.01852](https://doi.org/10.48550/arXiv.1502.01852), [arXiv:1502.01852](https://arxiv.org/abs/1502.01852).
- He, Z., Er, X., Long, Q., et al., 2020. Deep learning for strong lensing search: tests of the convolutional neural networks and new candidates from KiDS DR3. *MNRAS* 497, 556–571. doi:[10.1093/mnras/staa1917](https://doi.org/10.1093/mnras/staa1917), [arXiv:2007.00188](https://arxiv.org/abs/2007.00188).
- Hezaveh, Y., Dalal, N., Holder, G., et al., 2013. Dark Matter Substructure Detection Using Spatially Resolved Spectroscopy of Lensed Dusty Galaxies. *Astrophysical Journal* 767, 9. doi:[10.1088/0004-637X/767/1/9](https://doi.org/10.1088/0004-637X/767/1/9), [arXiv:1210.4562](https://arxiv.org/abs/1210.4562).
- Hildebrandt, H., Viola, M., Heymans, C., et al., 2017a. KiDS-450: cosmological parameter constraints from tomographic weak gravitational lensing. *MNRAS* 465, 1454–1498. doi:[10.1093/mnras/stw2805](https://doi.org/10.1093/mnras/stw2805), [arXiv:1606.05338](https://arxiv.org/abs/1606.05338).
- Hildebrandt, H., Viola, M., Heymans, C., et al., 2017b. KiDS-450: cosmological parameter constraints from tomographic weak gravitational lensing. *MNRAS* 465, 1454–1498. doi:[10.1093/mnras/stw2805](https://doi.org/10.1093/mnras/stw2805), [arXiv:1606.05338](https://arxiv.org/abs/1606.05338).
- Hinton, G.E., Osindero, S., Teh, Y.W., 2006. A fast learning algorithm for deep belief nets. *Neural Computation* 18, 1527–1554. URL: <https://api.semanticscholar.org/CorpusID:2309950>.

- Hoekstra, H., Franx, M., Kuijken, K., Squires, G., 1998. Weak Lensing Analysis of CL 1358+62 Using Hubble Space Telescope Observations. *Astrophysical Journal* 504, 636–660. doi:[10.1086/306102](https://doi.org/10.1086/306102).
- Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. *Neural Networks* 2, 359–366. URL: <https://www.sciencedirect.com/science/article/pii/0893608089900208>, doi:[https://doi.org/10.1016/0893-6080\(89\)90020-8](https://doi.org/10.1016/0893-6080(89)90020-8).
- Hu, W., 1998. Structure Formation with Generalized Dark Matter. *Astrophysical Journal* 506, 485–494. doi:[10.1086/306274](https://doi.org/10.1086/306274), [arXiv:astro-ph/9801234](https://arxiv.org/abs/astro-ph/9801234).
- Impellizzeri, C.M.V., McKean, J.P., Castangia, P., et al., 2008. A gravitationally lensed water maser in the early Universe. *Nature* 456, 927–929. doi:[10.1038/nature07544](https://doi.org/10.1038/nature07544), [arXiv:0901.1132](https://arxiv.org/abs/0901.1132).
- Ivezić, Ž., Kahn, S.M., Tyson, J.A., et al., 2019. LSST: From Science Drivers to Reference Design and Anticipated Data Products. *Astrophysical Journal* 873, 111. doi:[10.3847/1538-4357/ab042c](https://doi.org/10.3847/1538-4357/ab042c), [arXiv:0805.2366](https://arxiv.org/abs/0805.2366).
- Jacobs, C., Glazebrook, K., Collett, T., et al., 2017. Finding strong lenses in CFHTLS using convolutional neural networks. *MNRAS* 471, 167–181. doi:[10.1093/mnras/stx1492](https://doi.org/10.1093/mnras/stx1492), [arXiv:1704.02744](https://arxiv.org/abs/1704.02744).
- Jankowski, N., Duch, W., Grabczewski, K., 2013. Meta-Learning in Computational Intelligence. URL: <https://api.semanticscholar.org/CorpusID:263897152>.
- Jarrett, K., Kavukcuoglu, K., Ranzato, M., LeCun, Y., 2009. What is the best multi-stage architecture for object recognition?, in: 2009 IEEE 12th international conference on computer vision, IEEE. pp. 2146–2153.
- Kaiser, N., Squires, G., Broadhurst, T., 1995. A Method for Weak Lensing Observations. *Astrophysical Journal* 449, 460. doi:[10.1086/176071](https://doi.org/10.1086/176071), [arXiv:astro-ph/9411005](https://arxiv.org/abs/astro-ph/9411005).
- Keeton, C., 2014. Principles of Astrophysics. doi:[10.1007/978-1-4614-9236-8](https://doi.org/10.1007/978-1-4614-9236-8).
- Keeton, C.R., Kochanek, C.S., Seljak, U., 1997. Shear and Ellipticity in Gravitational Lenses. *Astrophysical Journal* 482, 604–620. doi:[10.1086/304172](https://doi.org/10.1086/304172), [arXiv:astro-ph/9610163](https://arxiv.org/abs/astro-ph/9610163).
- Kelly, P.L., Diego, J.M., Rodney, S., et al., 2018. Extreme magnification of an individual star at redshift 1.5 by a galaxy-cluster lens. *Nature Astronomy* 2, 334–342. doi:[10.1038/s41550-018-0430-3](https://doi.org/10.1038/s41550-018-0430-3), [arXiv:1706.10279](https://arxiv.org/abs/1706.10279).
- Kilbinger, M., 2015. Cosmology with cosmic shear observations: a review. *Reports on Progress in Physics* 78, 086901. doi:[10.1088/0034-4885/78/8/086901](https://doi.org/10.1088/0034-4885/78/8/086901), [arXiv:1411.0115](https://arxiv.org/abs/1411.0115).
- King, I.R., 1971. The Profile of a Star Image. *PASP* 83, 199. doi:[10.1086/129100](https://doi.org/10.1086/129100).
- Kochanek, C.S., 1996. The Flat-Spectrum Radio Luminosity Function, Gravitational Lensing, Galaxy Ellipticities, and Cosmology. *Astrophysical Journal* 473, 595. doi:[10.1086/178175](https://doi.org/10.1086/178175), [arXiv:astro-ph/9611231](https://arxiv.org/abs/astro-ph/9611231).
- Koopmans, L.V.E., 2005. Gravitational imaging of cold dark matter substructures. *MNRAS* 363, 1136–1144. doi:[10.1111/j.1365-2966.2005.09523.x](https://doi.org/10.1111/j.1365-2966.2005.09523.x), [arXiv:astro-ph/0501324](https://arxiv.org/abs/astro-ph/0501324).
- Koopmans, L.V.E., Bolton, A., Treu, T., et al., 2009. The Structure and Dynamics of Massive Early-Type Galaxies: On Homology, Isothermality, and Isotropy Inside One Effective Radius. *ApJ* 703, L51–L54. doi:[10.1088/0004-637X/703/1/L51](https://doi.org/10.1088/0004-637X/703/1/L51), [arXiv:0906.1349](https://arxiv.org/abs/0906.1349).

- Kormann, R., Schneider, P., Bartelmann, M., 1994. Isothermal elliptical gravitational lens models. *A&A* 284, 285–299.
- Kormendy, J., 1977. Brightness distributions in compact and normal galaxies. III. Decomposition of observed profiles in spheroid and disk components. *Astrophysical Journal* 217, 406–419. doi:[10.1086/155589](https://doi.org/10.1086/155589).
- Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems* 25.
- Kron, R.G., 1980. Photometry of a complete sample of faint galaxies. *ApJS* 43, 305–325. doi:[10.1086/190669](https://doi.org/10.1086/190669).
- Kuijken, K., 2011. OmegaCAM: ESO’s Newest Imager.
- Kuijken, K., Heymans, C., Dvornik, A., et al., 2019. The fourth data release of the Kilo-Degree Survey: ugrI imaging and nine-band optical-IR photometry over 1000 square degrees. *A&A* 625, A2. doi:[10.1051/0004-6361/201834918](https://doi.org/10.1051/0004-6361/201834918), [arXiv:1902.11265](https://arxiv.org/abs/1902.11265).
- Kuijken, K., Heymans, C., Hildebrandt, H., et al., 2015. Gravitational lensing analysis of the Kilo-Degree Survey. *MNRAS* 454, 3500–3532. doi:[10.1093/mnras/stv2140](https://doi.org/10.1093/mnras/stv2140), [arXiv:1507.00738](https://arxiv.org/abs/1507.00738).
- La Barbera, F., de Carvalho, R.R., Kohl-Moreira, J.L., et al., 2008. 2DPHOT: A Multi-Purpose Environment for the Two-Dimensional Analysis of Wide-Field Images. *PASP* 120, 681. doi:[10.1086/588614](https://doi.org/10.1086/588614), [arXiv:0804.1175](https://arxiv.org/abs/0804.1175).
- Laureijs, R., Amiaux, J., Arduini, S., et al., 2011. Euclid Definition Study Report. *arXiv e-prints*, [arXiv:1110.3193](https://arxiv.org/abs/1110.3193)doi:[10.48550/arXiv.1110.3193](https://doi.org/10.48550/arXiv.1110.3193), [arXiv:1110.3193](https://arxiv.org/abs/1110.3193).
- LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature* 521, 436–444. doi:[10.1038/nature14539](https://doi.org/10.1038/nature14539).
- LeCun, Y., Boser, B., Denker, J., et al., 1989a. Handwritten digit recognition with a back-propagation network. *Advances in neural information processing systems* 2.
- LeCun, Y., Boser, B., Denker, J., et al., 1989b. Backpropagation applied to handwritten zip code recognition. *Neural computation* 1, 541–551.
- LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86, 2278–2324.
- Li, R., Napolitano, N.R., Spiniello, C., et al., 2021. High-quality Strong Lens Candidates in the Final Kilo-Degree Survey Footprint. *Astrophysical Journal* 923, 16. doi:[10.3847/1538-4357/ac2df0](https://doi.org/10.3847/1538-4357/ac2df0), [arXiv:2110.01905](https://arxiv.org/abs/2110.01905).
- Li, R., Napolitano, N.R., Tortora, C., et al., 2020. New High-quality Strong Lens Candidates with Deep Learning in the Kilo-Degree Survey. *Astrophysical Journal* 899, 30. doi:[10.3847/1538-4357/ab9dfa](https://doi.org/10.3847/1538-4357/ab9dfa), [arXiv:2004.02715](https://arxiv.org/abs/2004.02715).
- Longo, G., Merényi, E., Tiño, P., 2019. Foreword to the Focus Issue on Machine Intelligence in Astronomy and Astrophysics. *PASP* 131, 100101. doi:[10.1088/1538-3873/ab2743](https://doi.org/10.1088/1538-3873/ab2743), [arXiv:1906.08349](https://arxiv.org/abs/1906.08349).
- Luppino, G.A., Kaiser, N., 1997. Detection of Weak Lensing by a Cluster of Galaxies at $z = 0.83$. *Astrophysical Journal* 475, 20–28. doi:[10.1086/303508](https://doi.org/10.1086/303508), [arXiv:astro-ph/9601194](https://arxiv.org/abs/astro-ph/9601194).

- Mahabal, A., Rebbapragada, U., Walters, R., et al., 2019. Machine Learning for the Zwicky Transient Facility. *PASP* 131, 038002. doi:[10.1088/1538-3873/aaf3fa](https://doi.org/10.1088/1538-3873/aaf3fa), [arXiv:1902.01936](https://arxiv.org/abs/1902.01936).
- Mahlke, M., Bouy, H., Altieri, B., et al., 2018. Mining the Kilo-Degree Survey for solar system objects. *A&A* 610, A21. doi:[10.1051/0004-6361/201730924](https://doi.org/10.1051/0004-6361/201730924), [arXiv:1711.02780](https://arxiv.org/abs/1711.02780).
- Marshall, P., Sandford, C., More, A., Buddelmeijer, H., 2015. HumVI: Human Viewable Image creation .
- Mellier, Y., 1999. Probing the Universe with Weak Lensing. *ARA&A* 37, 127–189. doi:[10.1146/annurev.astro.37.1.127](https://doi.org/10.1146/annurev.astro.37.1.127), [arXiv:astro-ph/9812172](https://arxiv.org/abs/astro-ph/9812172).
- Moffat, A.F.J., 1969. A Theoretical Investigation of Focal Stellar Images in the Photographic Emulsion and Application to Photographic Photometry. *A&A* 3, 455.
- Möller, O., Kitzbichler, M., Natarajan, P., 2007. Strong lensing statistics in large, $z < 0.2$, surveys: bias in the lens galaxy population. *MNRAS* 379, 1195–1208. doi:[10.1111/j.1365-2966.2007.12004.x](https://doi.org/10.1111/j.1365-2966.2007.12004.x), [arXiv:astro-ph/0607032](https://arxiv.org/abs/astro-ph/0607032).
- Newton, I., 1704. Opticks. URL: <https://archive.org/details/1704-newton-opticks-1ed/page/n1/mode/2up>.
- Oguri, M., 2006. The image separation distribution of strong lenses: halo versus sub-halo populations. *MNRAS* 367, 1241–1250. doi:[10.1111/j.1365-2966.2006.10043.x](https://doi.org/10.1111/j.1365-2966.2006.10043.x), [arXiv:astro-ph/0508528](https://arxiv.org/abs/astro-ph/0508528).
- Pan, S.J., Yang, Q., 2009. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering* 22, 1345–1359.
- Petrillo, C.E., Tortora, C., Chatterjee, S., et al., 2019a. Testing convolutional neural networks for finding strong gravitational lenses in KiDS. *MNRAS* 482, 807–820. doi:[10.1093/mnras/sty2683](https://doi.org/10.1093/mnras/sty2683), [arXiv:1807.04764](https://arxiv.org/abs/1807.04764).
- Petrillo, C.E., Tortora, C., Chatterjee, S., et al., 2017. Finding strong gravitational lenses in the Kilo Degree Survey with Convolutional Neural Networks. *MNRAS* 472, 1129–1150. doi:[10.1093/mnras/stx2052](https://doi.org/10.1093/mnras/stx2052), [arXiv:1702.07675](https://arxiv.org/abs/1702.07675).
- Petrillo, C.E., Tortora, C., Vernardos, G., et al., 2019b. LinKS: discovering galaxy-scale strong lenses in the Kilo-Degree Survey using convolutional neural networks. *MNRAS* 484, 3879–3896. doi:[10.1093/mnras/stz189](https://doi.org/10.1093/mnras/stz189), [arXiv:1812.03168](https://arxiv.org/abs/1812.03168).
- Pourrahmani, M., Nayyeri, H., Cooray, A., 2018. LensFlow: A Convolutional Neural Network in Search of Strong Gravitational Lenses. *Astrophysical Journal* 856, 68. doi:[10.3847/1538-4357/aaae6a](https://doi.org/10.3847/1538-4357/aaae6a), [arXiv:1705.05857](https://arxiv.org/abs/1705.05857).
- Protopapas, P., Giammarco, J.M., Faccioli, L., et al., 2006. Finding outlier light curves in catalogues of periodic variable stars. *MNRAS* 369, 677–696. doi:[10.1111/j.1365-2966.2006.10327.x](https://doi.org/10.1111/j.1365-2966.2006.10327.x), [arXiv:astro-ph/0505495](https://arxiv.org/abs/astro-ph/0505495).
- Razim, O., Cavuoti, S., Brescia, M., et al., 2021. Improving the reliability of photometric redshift with machine learning. *MNRAS* 507, 5034–5052. doi:[10.1093/mnras/stab2334](https://doi.org/10.1093/mnras/stab2334), [arXiv:2108.04784](https://arxiv.org/abs/2108.04784).
- Refsdal, S., 1964. On the possibility of determining hubble’s parameter and the masses of galaxies from the gravitational lens effect. *Monthly Notices of the Royal Astronomical Society* 128, 307–310.

- Richard, J., Kneib, J.P., Limousin, M., et al., 2010. Abell 370 revisited: refurbished Hubble imaging of the first strong lensing cluster. *MNRAS* 402, L44–L48. doi:[10.1111/j.1745-3933.2009.00796.x](https://doi.org/10.1111/j.1745-3933.2009.00796.x), [arXiv:0910.5553](https://arxiv.org/abs/0910.5553).
- Rosenblatt, F., 1963. Principles of neurodynamics. perceptrons and the theory of brain mechanisms. *American Journal of Psychology* 76, 705. URL: <https://api.semanticscholar.org/CorpusID:62710001>.
- Roy, N., Napolitano, N.R., La Barbera, F., et al., 2018. Evolution of galaxy size-stellar mass relation from the Kilo-Degree Survey. *MNRAS* 480, 1057–1080. doi:[10.1093/mnras/sty1917](https://doi.org/10.1093/mnras/sty1917), [arXiv:1807.06085](https://arxiv.org/abs/1807.06085).
- Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. *Nature* 323, 533–536. doi:[10.1038/323533a0](https://doi.org/10.1038/323533a0).
- Schipani, P., Capaccioli, M., Arcidiacono, C., et al., 2012. VST: from commissioning to science, in: Stepp, L.M., Gilmozzi, R., Hall, H.J. (Eds.), *Ground-based and Airborne Telescopes IV*, p. 84441C. doi:[10.1117/12.925481](https://doi.org/10.1117/12.925481).
- Schneider, P., Ehlers, J., Falco, E., 1992. *Gravitational lenses* springer-verlag. Berlin Inc., New York .
- Sérsic, J.L., 1963. Influence of the atmospheric and instrumental dispersion on the brightness distribution in a galaxy. *Boletín de la Asociación Argentina de Astronomía La Plata Argentina* 6, 41–43.
- Sersic, J.L., 1968. *Atlas de Galaxias Australes*.
- Shajib, A.J., Treu, T., Birrer, S., Sonnenfeld, A., 2021. Dark matter haloes of massive elliptical galaxies at $z \sim 0.2$ are well described by the Navarro-Frenk-White profile. *MNRAS* 503, 2380–2405. doi:[10.1093/mnras/stab536](https://doi.org/10.1093/mnras/stab536), [arXiv:2008.11724](https://arxiv.org/abs/2008.11724).
- Shajib, A.J., Vernardos, G., Collett, T.E., et al., 2022. Strong Lensing by Galaxies. *arXiv e-prints* , [arXiv:2210.10790](https://arxiv.org/abs/2210.10790)doi:[10.48550/arXiv.2210.10790](https://doi.org/10.48550/arXiv.2210.10790), [arXiv:2210.10790](https://arxiv.org/abs/2210.10790).
- Shan, H., Liu, X., Hildebrandt, H., et al., 2018. KiDS-450: cosmological constraints from weak lensing peak statistics - I. Inference from analytical prediction of high signal-to-noise ratio convergence peaks. *MNRAS* 474, 1116–1134. doi:[10.1093/mnras/stx2837](https://doi.org/10.1093/mnras/stx2837), [arXiv:1709.07651](https://arxiv.org/abs/1709.07651).
- Shu, Y., Brownstein, J.R., Bolton, A.S., et al., 2017. The Sloan Lens ACS Survey. XIII. Discovery of 40 New Galaxy-scale Strong Lenses. *Astrophysical Journal* 851, 48. doi:[10.3847/1538-4357/aa9794](https://doi.org/10.3847/1538-4357/aa9794), [arXiv:1711.00072](https://arxiv.org/abs/1711.00072).
- Soldner, J., 1801. *Astronomisches Jahrbuch für das Jahr ...: nebst einer sammlung de neuesten in die astronomischen wissenschaften einschlagenden abhandlungen, beobachten und nachrichten. f. dümmler*. Berliner Astronomisches Jahrbuch .
- Spiniello, C., Agnello, A., Napolitano, N.R., et al., 2018. KiDS-SQuaD: The KiDS Strongly lensed Quasar Detection project. *MNRAS* 480, 1163–1173. doi:[10.1093/mnras/sty1923](https://doi.org/10.1093/mnras/sty1923), [arXiv:1805.12436](https://arxiv.org/abs/1805.12436).
- Springel, V., White, S.D., Jenkins, A., et al., 2005. Simulations of the formation, evolution and clustering of galaxies and quasars. *nature* 435, 629–636.
- Swinbank, A.M., Webb, T.M., Richard, J., et al., 2009. A spatially resolved map of the kinematics, star formation and stellar mass assembly in a star-forming galaxy at $z = 4.9$. *MNRAS* 400, 1121–1131. doi:[10.1111/j.1365-2966.2009.15617.x](https://doi.org/10.1111/j.1365-2966.2009.15617.x), [arXiv:0909.0111](https://arxiv.org/abs/0909.0111).

- Tan, M., Le, Q.V., 2019. Efficientnet: Rethinking model scaling for convolutional neural networks. ArXiv abs/1905.11946. URL: <https://api.semanticscholar.org/CorpusID:167217261>.
- Taye, M.M., 2023. Theoretical understanding of convolutional neural network: Concepts, architectures, applications, future directions. *Computation* 11. URL: <https://www.mdpi.com/2079-3197/11/3/52>, doi:10.3390/computation11030052.
- Tortora, C., Napolitano, N.R., Romanowsky, A.J., Jetzer, P., 2010. Central Dark Matter Trends in Early-type Galaxies from Strong Lensing, Dynamics, and Stellar Populations. *ApJ* 721, L1–L5. doi:10.1088/2041-8205/721/1/L1, arXiv:1007.3988.
- Tortora, C., Napolitano, N.R., Spavone, M., et al., 2018a. The first sample of spectroscopically confirmed ultra-compact massive galaxies in the Kilo Degree Survey. *MNRAS* 481, 4728–4752. doi:10.1093/mnras/sty2564, arXiv:1806.01307.
- Tortora, C., Napolitano, N.R., Spavone, M., et al., 2018b. The first sample of spectroscopically confirmed ultra-compact massive galaxies in the Kilo Degree Survey. *MNRAS* 481, 4728–4752. doi:10.1093/mnras/sty2564, arXiv:1806.01307.
- Treu, T., 2010. Strong Lensing by Galaxies. *ARA&A* 48, 87–125. doi:10.1146/annurev-astro-081309-130924, arXiv:1003.5567.
- Treu, T., Koopmans, L.V.E., 2004. Massive Dark Matter Halos and Evolution of Early-Type Galaxies to $z \sim 1$. *Astrophysical Journal* 611, 739–760. doi:10.1086/422245, arXiv:astro-ph/0401373.
- Treu, T., Schmidt, K.B., Brammer, G.B., et al., 2015. The Grism Lens-Amplified Survey from Space (GLASS). I. Survey Overview and First Data Release. *Astrophysical Journal* 812, 114. doi:10.1088/0004-637X/812/2/114, arXiv:1509.00475.
- Trujillo, I., Aguerri, J.A.L., Cepa, J., Gutiérrez, C.M., 2001. The effects of seeing on Sérsic profiles - II. The Moffat PSF. *MNRAS* 328, 977–985. doi:10.1046/j.1365-8711.2001.04937.x, arXiv:astro-ph/0109067.
- Trujillo, I., Conselice, C.J., Bundy, K., et al., 2007. Strong size evolution of the most massive galaxies since $z \sim 2$. *MNRAS* 382, 109–120. doi:10.1111/j.1365-2966.2007.12388.x, arXiv:0709.0621.
- Turner, E.L., Ostriker, J.P., Gott, J. R., I., 1984. The statistics of gravitational lenses : the distributions of image angular separations and lens redshifts. *Astrophysical Journal* 284, 1–22. doi:10.1086/162379.
- van Uitert, E., Hoekstra, H., Joachimi, B., et al., 2017. Halo ellipticity of GAMA galaxy groups from KiDS weak lensing. *MNRAS* 467, 4131–4149. doi:10.1093/mnras/stx344, arXiv:1610.04226.
- Vanderplas, J., Connolly, A., 2009. Reducing the Dimensionality of Data: Locally Linear Embedding of Sloan Galaxy Spectra. *AJ* 138, 1365–1379. doi:10.1088/0004-6256/138/5/1365, arXiv:0907.2238.
- Vegetti, S., Koopmans, L.V.E., Auger, M.W., et al., 2014. Inference of the cold dark matter substructure mass function at $z = 0.2$ using strong gravitational lenses. *MNRAS* 442, 2017–2035. doi:10.1093/mnras/stu943, arXiv:1405.3666.
- Verde, L., Treu, T., Riess, A.G., 2019. Tensions between the early and late Universe. *Nature Astronomy* 3, 891–895. doi:10.1038/s41550-019-0902-0, arXiv:1907.10625.