IIPI

Experimental Physics II

Bachelor Thesis

Design of a CNN to establish location and mass of dark matter distribution

by means of weak gravity lensing

Written by

William Roster

Justus-Liebig University &
Technische Hochschule Mittelhessen

JUSTUS-LIEBIG UNIVERSITY GIESSEN
& TECHNISCHE HOCHSCHULE MITTELHESSEN
- UNIVERSITY OF APPLIED SCIENCES

II. PHYSICS INSTITUTE (SUBATOMIC PARTICLE PHYSICS)
GROUP : PD DR. JENS SÖREN LANGE
FB07

Design of a convolutional neural network to establish location and mass of dark matter distribution by means of weak gravity lensing

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
BACHELOR OF SCIENCE (B.SC.)

WILLIAM THOMAS ROSTER

Advisor: PD Dr. Jens Sören Lange
2nd reader: PD Dr. Sophia Heinz

Abstract

In this thesis, a convolutional neural network was trained in the artificial intelligence field of deep learning, in order to detect patterns created by the cosmological effect known as gravitational lensing. By doing so, the network is capable of identifying and locating dark matter distributions, often found to be concentrated within galaxy clusters, in deep field images.

Even though dark matter makes up 85% of the total mass in the universe, it only interacts with other particles via gravity. Thus laborious lensing and extensive X-ray surveys are necessary to successfully detect any dark matter accumulation. A trained neural network however, solves this problem by recognizing even the faintest of structures, while working through large data sets in little time. The data on which the network is trained and partially validated, is randomly generated as plots, programmed in python. These illustrations are then dismantled into their individual pixel values and later imported to the neural network. Based on Keras, the network uses multiple layers of filters to spot reoccurring arrangements of background galaxies.

It appears that the network performs extremely effective for well lit images, as the binary inputs hold more lensing features this way. Tasks such as object localization and regression were found to directly depend on the quality of the pre-processed data.

Keywords

Contents

Chapter 1

Introduction

Mankind has always been a curious species. Humans themselves however, no more than clumps of atoms present for only an instant in eternity. And yet generations of humans have persistently been driven by a pioneering spirit sewn deep into their DNA, to gain knowledge, whilst filling in the blanks. It comes at no surprise, that at the forefront of man's greatest intrigue has and always will be the night sky and all the hidden worlds within, that are only ever revealed when darkness falls upon our horizon. Captured by the sheer magnitude and mystique of the universe some of the greatest intellectuals to have ever lived dedicated their lives towards determining our role in the spectrum in relation to what lies beyond.

An indispensable tool for any researcher who tried to further resolve the laws of nature bestowed upon us is the natural science of physics. But simultaneously our ever advancing understanding has consistently been limited to its technological boundaries. Thus scientific progress goes hand in hand with technological improvement, as we continue craving to explore worlds either too small or too large for humans to comprehend in an analytical manor.

A giant leap forward was made in the mid 1940's as John W. Mauchly and J. Priester Eckert built the first substantial computer at the University of Pennsylvania called ENIAC (Electrical Numerical Integrator and Calculator). Since then computers have continuously been improved and so has our research in regions which we would have never dreamed of exploring a few decades ago. Today we have reached a point where life without computers would be unimaginable. From the very basis at which the internet allows humans to connect, all the way up to high performance computing which we use to solve complex mathematical equations. Whilst doing so we are further lifting the curtain of ignorance.

Even though manmade machines such as computers help us immensely in looking for answers to solve the mysteries of the universe, it is not them who are asking the questions. After all, it takes skills such as perception, reasoning, learning and cognitive response to be able to have what you might call thoughts or intelligence in general. Fortunately for us the biological neural network humans use to perform learning tasks on a daily basis can be portrayed in a simplistic manor with values of 1 and 0 at its very basis. This circumstance serves as the corner stone of the implementation of so-called AI (Artificial Intelligence).

The applications for artificial intelligence are endless as it can be applied to almost all sectors and industries. As technology advances, machines are wired using a cross-disciplinary approach based on mathematics, computer science, linguistics, psychology, and more. Therefore previous benchmarks which defined artificial intelligence become outdated. Examples of machines with artificial intelligence include computers which play chess and self-driving-cars. Each of these machines must weigh the consequences of any action they take, as each action will impact upon the end result. In chess, the end result is winning the game. For self-driving cars, the computer system must account for all external data and compute it to act in a way that prevents a collision. Naturally scientists have taken to AI as it allows for a whole cluster of new methods on uncovering even the most remote and inaccessible data sets to date.

Apart from making new discoveries with this technology many researchers are also drawn to lifting questions that remain unresolved thus far. One example is dark matter, or "matière obscure" in French, as called by Henri Poincaré in 1906. The late 19th century already features first talks on dark bodies in the milky way derived from the orbital mechanics of stars circling the centre of our galaxy. Without the necessary technological advances at hand throughout the 20th century, dark matter has mostly remained a mystery. As the picture of dark matter becomes clearer with time, humanity hopes to further understand its complexity and put it in place with the other wonders of the universe.

At the beginning of the millennium more insight regarding dark matter was won, when examining the galaxy cluster 1E 0657- 56, also known as the "Bullet Cluster". This cluster was formed after the collision of two large clusters of galaxies, the most energetic event known in the Universe since the Big Bang. The concentration of mass is determined using the effect of so-called gravitational lensing, where by light from distant objects is distorted by intervening matter. After studying visual images of the bullet cluster and calculating the space curvature based on Newtons laws of gravity direct evidence was given that nearly all of the matter in the clusters is dark.

This thesis is devoted to training an artificial intelligence system by means of image recognition with diffraction patterns. The aim is for the artificial neural network to make reliable estimations regarding location and mass of dark matter accumulation in space based on the input data alone. By doing so a first step can be taken towards tangible analysis of dark matter within our universe and thus ultimately resulting in a better overall understanding.

All necessary physical problems and fundamentals will be addressed at a commensurate level at the beginning of the thesis. Afterwards the set up and implementation as well as the mode of operation of the AI will be explained thoroughly. Following the presentation of the experimental data concluding results will be evaluated. Lastly an outlook in regard of the recently acquired knowledge concerning dark matter distribution and its relation to the laws of theoretical physics will be projected.

Chapter 2

Physical basics

The physical problems addressed in this thesis are introduced in this chapter. First, the principles of gravity and relativity in the frame of the universe and the accompanying curvature of space-time are explained. Also, exploited effects such as light diffraction are discussed in the context of gravitational lensing.

2.1 Newtonian gravity

The beginning of modern mathematical physics can be dated back to the 5th July 1687, when Isaac Newton published the "Philosophiæ Naturalis Principia Mathematica" often referred to as simply the "Principia" [1]. Among other assumptions and axioms it comprises of his laws of motion and universal gravity. The latter is demonstrated to be a mathematical model which can be applied to explain or predict the behaviour of all objects in the universe by a single mathematical equation:

$$\vec{F}_{\mathrm{g}}(\vec{r}) = G \cdot \frac{m_1 \cdot m_2}{\vec{r^2}} \tag{2.1}$$

This equation obviously states that the gravitational force between two objects can be calculated by inserting their individual masses, relative distance to one another and then multiplying by G, which encodes the strength of the gravitational force itself. The so-called Newtonian constant is a fundamental property of the universe.

Today it can be measured at G = (6.6726 ±0.0005) x 10^{-8} $\frac{cm^3}{gs^2}$ which indirectly points out how weak of a force gravity is. This is partially the reason why G was first measured 71 years after Newton's death by Henry Cavendish in 1798.

Newton's laws are the key to understanding basic mechanics in the universe. They allow for the scale and geometry of the universe to be deduced. Also the orbital positions of celestial bodies can be calculated at any point in the future. Newton discovered his law of gravity by looking for a simple equation that could describe the apparent complexity of the motions of the planets around the sun. Keplers three empirical laws of planetary motion form the basis as they can be derived from Newton's law of gravity and his laws of motion. His was the first truly universal law of nature to be discovered.

Newtonian cosmology is based on several assumptions which are summarized in the cosmological principle:

- the universe shows a Euclidean structure

- gravity is the only force in the universe

- the universe is filled with matter uniformly to infinite distances

- the universe is consistently homogeneous and isotope

- no point is of preference

However this model does not add up mathematically if one assumes that space is infinitely and uniformly filled with bodies that attract each other as undefined present gravitational forces exercise considerable movement [1]. This instability indicates that Newton's law of gravitation can therefore only be applied if all matter is continuously in motion. This is also known as the gravitational paradox [1].

Based on this; Harlow Shapley published the first of a series of papers that refined the method used to calculate distances to the stars beyond our solar system established by Danish astronomer Ejnar Hertzsprung in 1915. Amongst other things this led him to the first measurement of the size and shape of the Milky Way. Today we know that we live on a disk of stars, spreading 100 000 light years across; with the sun 25 000 lightyears from its centre [2]. Great progress had been made. But from a physicists perspective there is no doubt that physics experienced a revolution in 1915, because in the November of that year Albert Einstein presented a new theory of gravity to the Prussian Academy of Science. The theory is known as General Relativity (GR) and it replaces Newton's law of universal gravitation. Many physicists regard GR as the most beautiful piece of physics yet devised by the human mind. The parallels are clear. Without Newton there would be no fundamental understanding of relating orbital mechanics, whilst without Einstein there would be no proper comprehension of the structure and behaviour of the universe.

2.2 General relativity

Einstein is most famous for his equation $E = mc^2$. The equivalence of mass and energy also clarifies the assumption that energy densities can be viewed as sources of gravitational fields. Since these are not linear, it is not surprising that both the special relativity theory published in 1905 and the differential equations of the GR are of non-linear character. At the heart of the theory is a very simple assumption that dates back all the way to Galileo. Put simply, there is no way one can tell whether one is moving or not. As long as there is no external force applied causing one to accelerate, the right of claiming to be "at rest" can be taken up at all times. However the all conquering Universal Law of Gravitation did not fit within the framework of special relativity. Therefore modifications were necessary. Described by Einstein as "glücklichster Gedanke meines Lebens", *the happiest thought of my life*, it occurred to him that the gravitational field has only a

---

[1]H.v. Seeliger, 1895

relative existence. In a way it is similar to the electric field generated by induction. As for an observer who is free falling, there exists no gravitational field at least in his immediate surroundings. With Newton's apple falling from a tree one might conclude that in Einsteins universe; Newton as well as the planet beneath his feet are conversely moving upwards to meet the falling apple [1]. Published in 1916 Einsteins theory of GR replaced the force of gravity with geometry and his four-dimensional space-time continuum of non-Euclidean form, also known as Riemannien geometry [2], in particular, the curvature of space and time.

The deviations of the Euclidean form are due to the distribution and movement of gravitational masses in space. The deviation from the Euclidean view is measured by the radius of curvature. The shortest connection between two points in space is then no longer just a straight line, but a so-called geodesic line, the course of which depends on the metric structure of the point under consideration. Two parallel lines on the surface of a sphere for example move ever closer as they continue to be projected. This is how geometry can lead to the appearance of a force. Einstein's theory of gravity allows one to calculate how space and time are curved by the presence of matter and energy and how objects move across the curved spacetime. Spacetime is often described as the fabric of the universe [3]. Massive objects such as stars and planets tell the fabric how to curve and the fabric tells the objects how to move. This is the GR equivalent of Newton's first law of motion - every body continues in a state of rest or uniform motion in a straight line unless acted upon by a force.
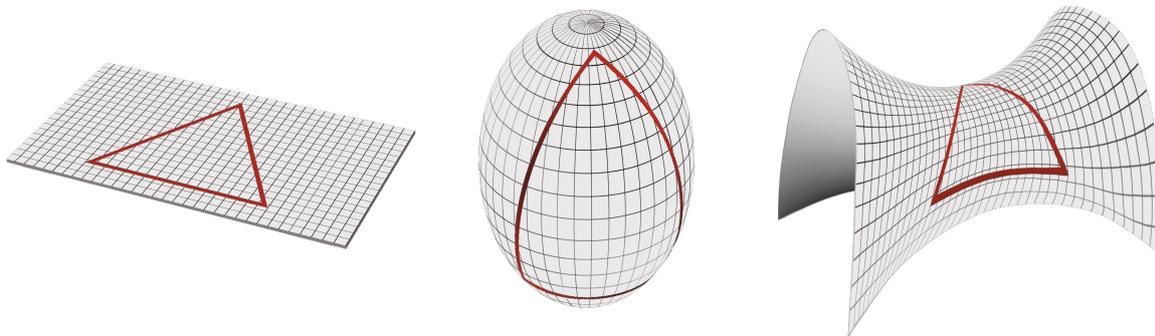

Figure 2.1: *Illustration of a straight line conversion to a geodesic for triangles with a deviating sum of angles for Euclidean, Spherical and Hyperbolical geometries. Image credit: Janus Cosmological*

Earth for example orbits around the sun is a straight line in spacetime curved by the presence of the sun. Naturally Earth continues to follow this straight line, because there is no external force to suggest an alternation. This is the opposite of the newtonian description, in which the earth were to move through space in a straight line, if it wasn't for the sun's gravitational force. The lines themselves appear curved to bystanders for the simple reason that the space upon which the straight lines are pictured is curved itself. However there is more to Einsteins GR than just the description of orbits. GR is fundamentally different to Newton's law because is doesn't simply provide a model for the action of gravity, rather, it provides an explanation for the existence of the gravitational force itself in terms of the curvature of space-time.

---

[2]Named after the german mathematician Bernhard Riemann

2.2.1 Einstein field equations

Newton's universal theory of gravitation determines the gravitational potential of a body with mass M to be:

$$\Phi(\vec{r}) = \frac{GM}{\vec{r}} \tag{2.2}$$

In general, the peculiarities of the general theory of relativity have to be taken into account if the gravitational energy becomes comparable to the self-energy [4]. Einstein's relativistic field theory thus includes Newton's theory of gravity as a borderline case. In order to clarify this better, Newton's law of motion, which is known to be

$$m_{\mathrm{i}} \frac{d^2 \vec{r_{\mathrm{i}}}}{dt^2} = -G \frac{m_{\mathrm{i}} m_{\mathrm{j}} (\vec{r_{\mathrm{i}}} - \vec{r_{\mathrm{j}}})}{|\vec{r_{\mathrm{i}}} - \vec{r_{\mathrm{j}}}|^3} \tag{2.3}$$

features conic sections as solutions of this equation of motion [5]. They can for example describe bound or unbound orbits. However, the speed of light is so large that it exceeds the escape velocity. Thus, the resulting orbit will be an hyperbolic orbit. Alternatively it can also be written in the form of

$$\nabla_{\mathrm{i}}^2 \Phi_{\mathrm{ij}}(\vec{r_{\mathrm{i}}}) = 4\pi G \rho_{\mathrm{j}}(\vec{r_{\mathrm{i}}}) \tag{2.4}$$

$$m_{\mathrm{i}} \frac{d^2 \vec{r_{\mathrm{i}}}}{dt^2} = -m_{\mathrm{i}} \nabla_{\mathrm{i}} \Phi_{\mathrm{ij}}(\vec{r_{\mathrm{i}}}) \tag{2.5}$$

where the gravitational potential results from the mass density used within the Poisson equation. In this way, the gravitational force is determined directly from the gravitational potential of the associated mass distribution. The relation regarding gravity made by Einstein is best described by his well known field equations, a system of ten coupled, nonlinear, hyperbolic-elliptical differential equations for the metric tensor $g_{\mu\nu}$. It can be seen as the relativistic generalization of Newton's gravitational potential:

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = \frac{G 8\pi}{c^4} T_{\mu\nu} \tag{2.6}$$

Formula without the cosmological constant term $\Lambda g_{\mu\nu}$ for reasons of simplicity

The right hand side describes the distribution of matter and energy in some region of space-time and the left hand side describes the shape of space-time as a result of the matter and energy distribution. With given matter distribution, Einsteins equations allow calculations as to what space-time looks like. Within the framework of GR, the equation of motion of a particle with the trajectory parameterized by "u" in the gravitational field, is formulated as

$$\frac{d^2 x^\lambda}{du^2} = -\Gamma_{\mu\nu}^{} \frac{dx^\mu}{du} \frac{dx^\nu}{du} \tag{2.7}$$

With:

- Christoffel-symbol $\Gamma_{\mu\nu}$

- Ricci-tensor $R_{\mu\nu}$

- Energy-impuls-tensor $T_{\mu\nu}$

In the context of this bachelor's thesis, the mathematical operators of these equations will not further be discussed. The borderline case for newtonian calculations is found to be at

$$g_{00} \approx -1 - \frac{2\Phi}{c^2} \tag{2.8}$$

In other words, Newton's approximation can only be used if $\frac{2\Phi}{c^2} \ll 1$, meaning that essentially it depends on the relationship between

$$\frac{2|\Phi|}{c^2} \approx \begin{cases} 10^{-4} & \text{earth} \\ 4 \cdot 10^{-6} & \text{sun} \\ 3 \cdot 10^{-4} & \text{white dwarf} \\ 0.3 & \text{neutron star} \end{cases} \tag{2.9}$$

Generally, a check of the potential Φ being much smaller than $\frac{c^2}{2}$ is sufficient. This clearly implicates that Newton's laws can be used for most calculations but denser objects create space curvature which can only be described in a relativistic manor. Central to the exploration of motion has been the idea of a reference frame.
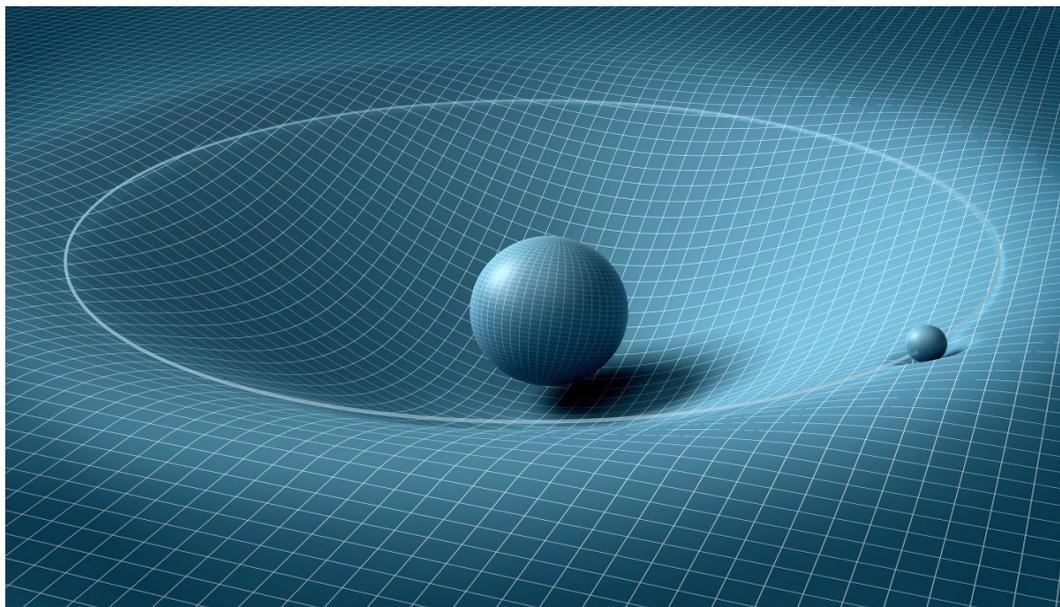

Figure 2.2: *The fabric of curved space exercising an apparent force on a mass. Credit: Now.de.*

The idea is that it is impossible to work out which inertial reference frame one is in, as they are all absolute equivalent. The statement that all inertial reference frames are equal is one of symmetry. Symmetry in maths and physics means doing something with the result staying the same. So when observing motion in an inertial reference frame of Newtons laws of gravity, and then changing the frame the symmetry in Newtons law allows for one to still make proper calculations. However the change of the reference system must be recognised. This means that Newton's laws are invariant under "Galileon Transformations" [6]. However this varies when the inertial system is changed into a rotating or accelerated reference system. Accompanied by fictitious forces the newtonian laws are now no longer invariant. According to the general principle of relativity, inertial forces are in principle indistinguishable from gravitational forces. This is also given by the equivalence of gravitational and inert masses. This is partly seen as the major reason why Einstein replaced Newton's law of gravity with his theory of GR. Even though Einstein's theory of GR has an array of important implications in astrophysical cosmology such as:

- the existence of black holes - borderline cases of space-time distortion - as an end state for massive stars

- the prediction of gravitational waves

- the expansion of the universe as one and, in contrary, its beginning, also known as Big Bang

This thesis will focus only on the diffraction of light due to gravity which can lead to a phenomenon known as gravitational lensing, in which multiple images of the same distant astronomical objects are visible in the sky.

2.3   Diffraction - the approximation of scalar waves

The effect given by the wave nature of light explained in the following, ensures that light deviates from its direction of movement as long as it spreads in a homogeneous and isotope medium when it passes an obstacle or through an opening [7]. As a result, light does not spread in a straight line, but can also be seen where, according to the laws of geometric optics, there should be a shadow. The so-called Huygens-Fresnel principle provides an explanation. Put simply, the waves intersect because each point of a wave front serves as a center of a new elementary wave.

- **Frenetic zones**

  According to a suggestion made by Fresnel in the early 19th century, all waves superimposed on a spatial point can be characterized as sources of concentric circles on a wavefront. The waves emanating from a Fresnel zone have path differences smaller than $\frac{\lambda}{2}$. This means that destructive interference does not occur when two waves superpose [8]. However, for each center of a zone there is a center either in the half adjoining outside or half adjoining inside in such a way that they extinguish each other with path differences at

exactly $\frac{\lambda}{2}$. An exception to this rule only applies to the first half and last half.
This ultimately results in only the maxima being visible as the remaining waves cancel
each other out. The area taken up by the zones can be calculated as:

$$A_{\mathrm{n}} = \pi n a \lambda \tag{2.10}$$

With the help of this formula, the first frenetic zone can be calculated - and from this the
maximum deflection of the light.

$$\theta = arctan \frac{1}{\sqrt{2}} \sqrt{a\lambda} \tag{2.11}$$

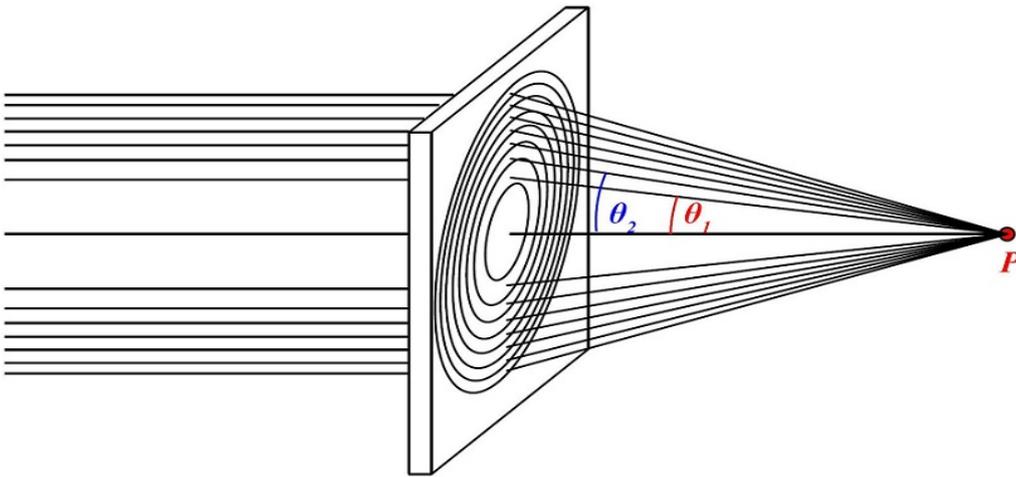

Figure 2.3: *Fresnel zones and reconstruction of a transmissionhologram. Credit: KIT.*

The Fresnel approach is transferred to Fraunhofer diffraction if the distance between the
exciter and the object is large compared to the size of the object. In this approach plane
waves meet the diffraction object. Therefore, only Fraunhofer diffraction will be considered
in the following.

2.3.1 Fermat's principle

As previously established, the consideration of curved spacetime and the use of geodesics need
only be used to correctly describe the present metric, if the Newtonian gravitational potential Φ
or the relative velocity v have large proportions in relation to c. Under such conditions,
gravitational lensing can be described by a small perturbation of the locally Minkowskian
space-time of an observer co-moving with the gravitational lens [12]. The Minkowski metric of
special relativity, expressed by its line element can be described by the simplified
Friedmann-Robertson-Walker metric:

$$ds^2 = \sum_{i=0}^{3} \sum_{j=0}^{3} dx_{\mathrm{i}} dx_{\mathrm{j}} g_{\mathrm{ij}} = -c^2 \left(1 + \frac{2\Phi}{c^2}\right) dt^2 + \left(1 - \frac{2\Phi}{c^2}\right) d\vec{x}^2 \tag{2.12}$$

It turns out that light deflection can equivalently be described by Fermat's principle, as in
geometrical optics.

Fermat's principle states that the path taken by a ray between two given points is the path that can be traversed in the least time. In order to be true in all cases, this statement must be weakened by replacing the "least" time with a time that is "stationary" with respect to variations of the path.
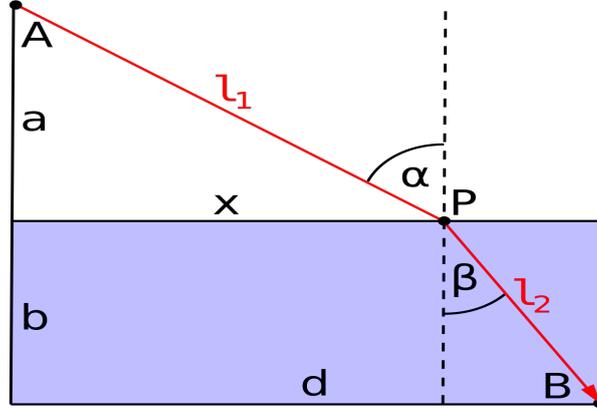

Figure 2.4: *Light refraction as perceived for medium crossing rays. Credit: Wikipedia.*

The speed of light in a medium with refractive index n is c/n, where c is its speed in a vacuum. Thus, the time required for light to go some distance in such a medium is n times the time light takes to go the same distance in a vacuum.

With the propagation condition for light, ds=0, this expression can be rearranged to find the effective light speed in a weak gravitational field

$$c' = \left| \frac{dx}{dt} \right| = c \left(1 + \frac{2\Phi}{c^2} \right) \tag{2.13}$$

Introducing the index of refraction n by the conventional definition c=c/n, we see that a weak gravitational field has the effective index of refraction

$$n = \frac{c'}{c} = 1 - \frac{2\Phi}{c^2} \tag{2.14}$$

As such a light ray geodesic ranging from two points A to B can be written as

$$\int_A^B n[\vec{x}(l)]dl \tag{2.15}$$

Furthermore n accounts for the eigentime ds to proportional resulting in the light path

$$\delta_\tau = \delta \int_A^B \frac{c}{n} dt = 0 \tag{2.16}$$

The variation of τ with respect to the light path leads to the deflection angle

$$\hat{\alpha} = \frac{2}{c^2} \int_{\lambda_A}^{\lambda_B} \vec{\nabla}_\perp \Phi d\lambda = \frac{4GM}{Rc^2} \tag{2.17}$$

which is the gradient of the dimension-less Newtonian potential perpendicular to the light ray, integrated along the light ray and multiplied by two. Since the speed of the light is reduced in the gravitational field, c = c/n, the travel time (along the perturbed path) increases.

This is the so-called Shapiro delay.[3]

$$\triangle t = \int \frac{dl}{c'} - \int \frac{dl}{c} = \int (n-1)dl = -\frac{2}{c^3} \int \Phi dl \tag{2.18}$$

2.4 Gravity lensing

As early as 1911, Einstein pursued his assumption of light deflection in the curvature of space caused by massive astronomical objects. He recognized, that due to the principle of equivalence, the trajectory of photons in a gravitational field must also be curved [10]. According to the well-known Einstein energy-mass relation, a mass can be assigned to the photon of energy $E = h\nu$. If photons possess a mass, they must also be able to be influenced by gravitational fields.

$$E = m_{\mathrm{ph}}c^2 = h\nu \qquad \Rightarrow \qquad m_{\mathrm{ph}} = \frac{h\nu}{c^2} \tag{2.19}$$

A deflection of photons moving on null-geodesics should be observable, especially when passing the sun. He published a version of what later would become the theory of the GR: "About the influence of gravity on the propagation of light[4]. In this manuscript he predicts a deflection of 0.85 arcseconds. In the final version in 1916, the value doubled to 1.75 arcseconds. Previously, the calculations had been made without the effects of the curvature of space and the shift in time for reference systems specific observers, causing twice the deflection to occur.
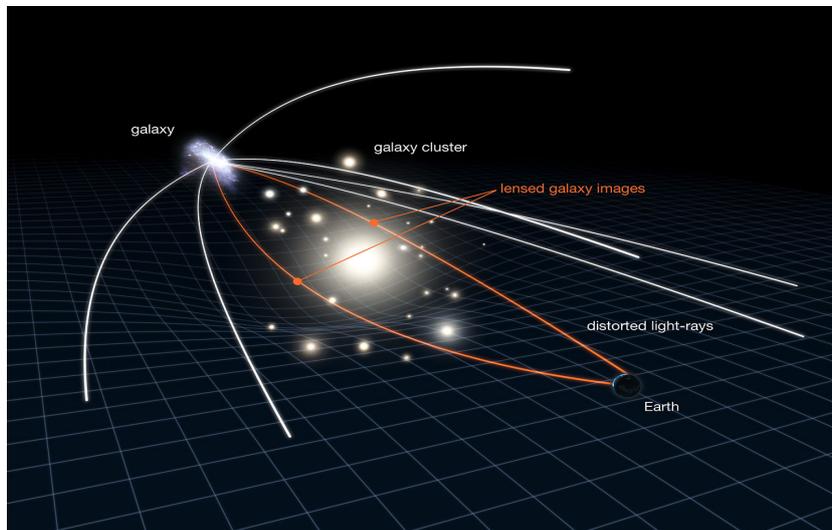

Figure 2.5: *Light diffraction as perceived on earth due to space distortion caused by large astronomical objects. Credit: Forbes.com*

It must be taken into account that the light also moves in time, so that there is actually a curvature of space-time and not a pure curvature of the three-dimensional space. For this purpose, the relative deviation of photons travelling at the speed of light with respect to the change in the gravitational potential, depending on the absolute distance, can be integrated over the angular space of 180°.

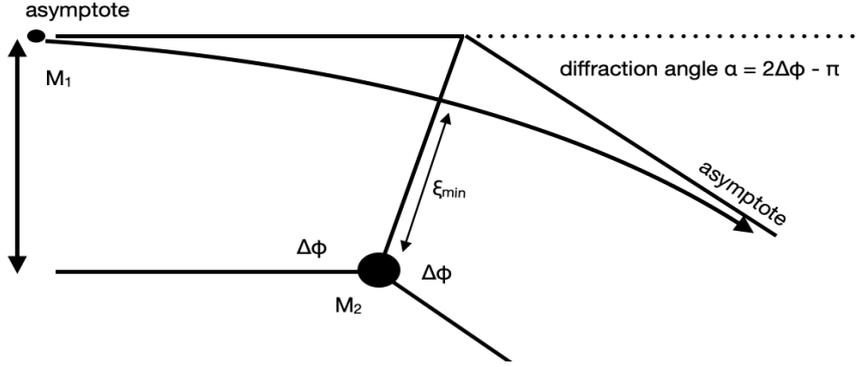[3]Shapiro, 1964
[4]Ann. Phys. 35 (1911), s. 895

Figure 2.6: *Illustration of geometric perception of the diffraction angle. Credit: [4]*

The factors of space curvature and time shift, which are summarized as γ, also influence this relationship [11]. For particles without resting-mass, γ was determined to have the value $\gamma = 1$. The final diffraction can be calculated using the Schwarzschild-radius r_s:

$$\hat{\alpha}(\xi_\mathrm{min}) = 2(\gamma + 1) \cdot \frac{GM}{c^2 \xi_\mathrm{min}} = \frac{2r_\mathrm{s}}{\xi_\mathrm{min}} \tag{2.20}$$

	mass($\mathrm{M_\odot}$)	size(pc)	$\alpha(arcsec)$
Sun	1	10^{-7}	1
Galaxy	10^{11}	10^{4}	10
Galaxy cluster	10^{14}	10^{5}	100

Table 2.1: *Comparison of angular diffraction for objects of different mass.*

This approach also provides an explanation for light deflection, which can be interpreted as a kind of refraction. These predictions were experimentally confirmed a few years later when the astronomers Eddington and Sobral compared photo plates with images of fixed stars during a solar eclipse on 29th May, 1919 with those previously recorded at night.
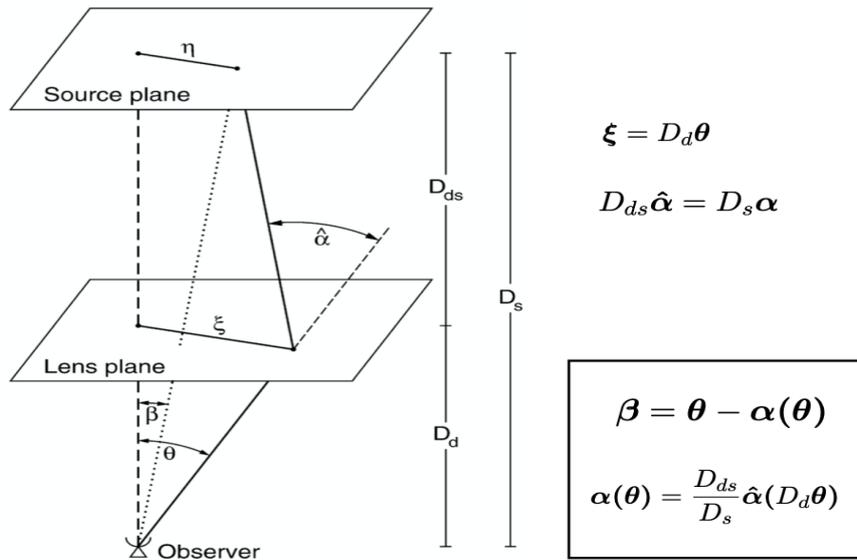

$$\boldsymbol{\xi} = D_d \boldsymbol{\theta}$$

$$D_{ds} \hat{\boldsymbol{\alpha}} = D_s \boldsymbol{\alpha}$$

$$\boldsymbol{\beta} = \boldsymbol{\theta} - \boldsymbol{\alpha}(\boldsymbol{\theta})$$

$$\boldsymbol{\alpha}(\boldsymbol{\theta}) = \frac{D_{ds}}{D_s} \hat{\boldsymbol{\alpha}}(D_d \boldsymbol{\theta})$$

Figure 2.7: *Scheme of a typical lensing system. Credit: Bartelmann & Schneider, 2001*

If $\vec{\theta}$, $\vec{\beta}$, and $\vec{\alpha}$ are small, the true and the observed position of the source are linked by a simple equation:

$$\vec{\theta}D_{\mathrm{s}} = \vec{\beta}D_{\mathrm{s}} + \vec{\alpha}D_{\mathrm{ds}} \qquad (2.21)$$

The apparent local distortion of a fixed star in the night sky, caused by the curvature of the light rays around an object rich in mass striking the observer, only includes a simplistic representation of what is otherwise known as gravitational lensing [12] [13] [14].
As can be seen in formula 2.13 a gravitational lens produces a maximum deflection of light that passes closest to its center, and a minimum deflection of light which travels furthest from its center. Therefore, unlike an optical lens, a gravitational lens has no single focal point, but a focal line. Solely depending on the gravitational potential at hand, lensing effects can be categorised into two main classes.

2.4.1 Strong lensing

If the source, the massive lensing object, and the observer lie in a straight line, the original light source will appear as a ring around the massive lensing object, also known as the *Einstein Ring*. More commonly, where the lensing mass is complex such as a galaxy cluster and does not cause a spherical distortion of spacetime, the source will resemble partial arcs scattered around the lens. The observer may then see multiple distorted images of the same source - also known as the *Einstein cross* -; the number and shape of these depending upon the relative positions of the source, lens, and observer. This relation, as can be seen in 2.9, can be derived from the lens equation

$$\beta = \theta - \theta_{\mathrm{E}}^2 \frac{\theta}{|\theta|^2} \qquad (2.22)$$

$$\alpha(\theta) = \frac{4GM}{c^2} \frac{D_{\mathrm{ds}}}{D_{\mathrm{d}}D_{\mathrm{s}}} \frac{\theta}{|\theta|^2} \qquad (2.23)$$

$$\theta_{\mathrm{E}} = \sqrt{\frac{4GM}{c^2} \frac{D_{\mathrm{ds}}}{D_{\mathrm{d}}D_{\mathrm{s}}}} \qquad (2.24)$$

Generally, the strong lensing effect requires the projected lens mass density to be greater than the critical density

$$\sum_{cr} = \frac{c^2}{4G\pi} \frac{D_{\mathrm{s}}}{D_{\mathrm{ds}}D_{\mathrm{d}}} = 0.35 \frac{g}{cm} \left(\frac{D_{\mathrm{ds}}D_{\mathrm{d}}}{D_{\mathrm{s}}} \frac{1}{Gpc} \right)^{-1} \qquad (2.25)$$

When multiple images appear, they tend to hover around the Einstein radius. This correlation can be used to estimate the mass of the galaxy or even a galaxy cluster since the mean surface mass density inside the Einstein radius equals the critical mean density [15].

$$\left\langle \sum (\theta_{\mathrm{arc}}) \right\rangle \approx \left\langle \sum (\theta_{\mathrm{E}}) \right\rangle = \sum_{cr} \qquad (2.26)$$

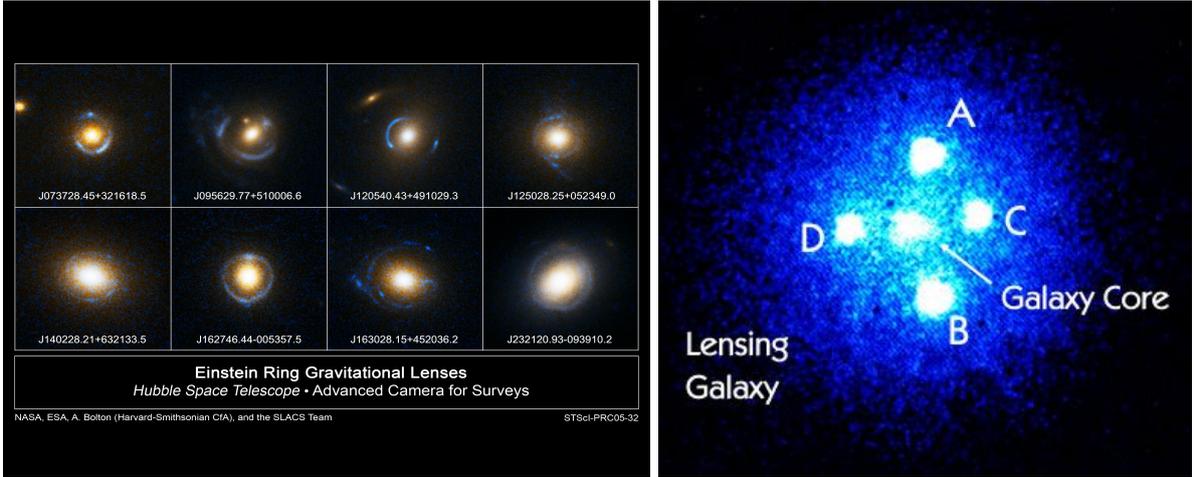$$M(\theta) = \sum_{cr} \pi (D_{\mathrm{d}}\theta)^2 \qquad (2.27)$$

Figure 2.8: *Einstein Ring captured by the Hubble Space Telescope. Credit: NASA*

Figure 2.9: *Einstein Cross caused by off centered light sources in the background. Credit: Wikipedia*

2.4.2 Weak lensing

Even though it is well established that any presence of mass bends the path of light, this effect does not portrait itself by producing giant arcs and other phenomena as associated with strong lensing. In fact, the majority of lensing incidents are within the weak lensing regime. Regardless of the fact that it is impossible to detect deflection in the order of 10% in a single background source, the presence of the foreground mass can be detected, by systematical alignment of background sources around the lensing mass. Thus this method is intrinsically statistical for preferred stretching of the background objects, perpendicular to the direction towards the centre of the lens. To measure this tangential alignment, it is necessary to measure the ellipticities of the background galaxies and construct a statistical estimate of their systematic alignment. The fundamental problem is that galaxies are not intrinsically circular, so their measured ellipticity is a combination of their intrinsic ellipticity and the gravitational lensing influence. The measurements of many background galaxies must be combined to average down this "shape noise". As the orientation of galaxies is considered to be entirely random, any systematic alignment between multiple galaxies can generally assumed to be caused by lensing. This coordinate transformation of the background objects can be split into two terms, the convergence and shear [16].

An extended distribution of matter is characterized by its effective lensing potential, obtained by projecting the three-dimensional Newtonian potential on the lens plane and by properly rescaling it:

$$\Psi(\vec{\theta}) = \frac{D_{\mathrm{ds}}}{D_{\mathrm{s}}D_{\mathrm{d}}} \frac{2}{c^2} \int \Phi(D_{\mathrm{d}}\vec{\theta}, z)dz \tag{2.28}$$

This lensing potential satisfies the two important properties of

$$\vec{\nabla}_\theta \Psi(\vec{\theta}) = \vec{\alpha}(\vec{\theta}) \tag{2.29}$$

$$\triangle_\theta \Psi(\vec{\theta}) = 2\kappa(\vec{\theta}) = \Phi_{11} + \Phi_{22} \tag{2.30}$$

14

with $\sum(\vec{\theta})$

$$\sum(\vec{\theta}) = \int \rho(\vec{\xi}, z)dz = \frac{1}{4G\pi} \int \triangle \Phi dz \tag{2.31}$$

and the convergence κ defined as:

$$\kappa(\vec{\theta}) = \frac{\sum(\vec{\theta})}{\sum_{cr}} = \frac{1}{c^2} \frac{D_{\mathrm{d}}D_{\mathrm{ds}}}{D_{\mathrm{s}}} \int \triangle \Phi dz \tag{2.32}$$
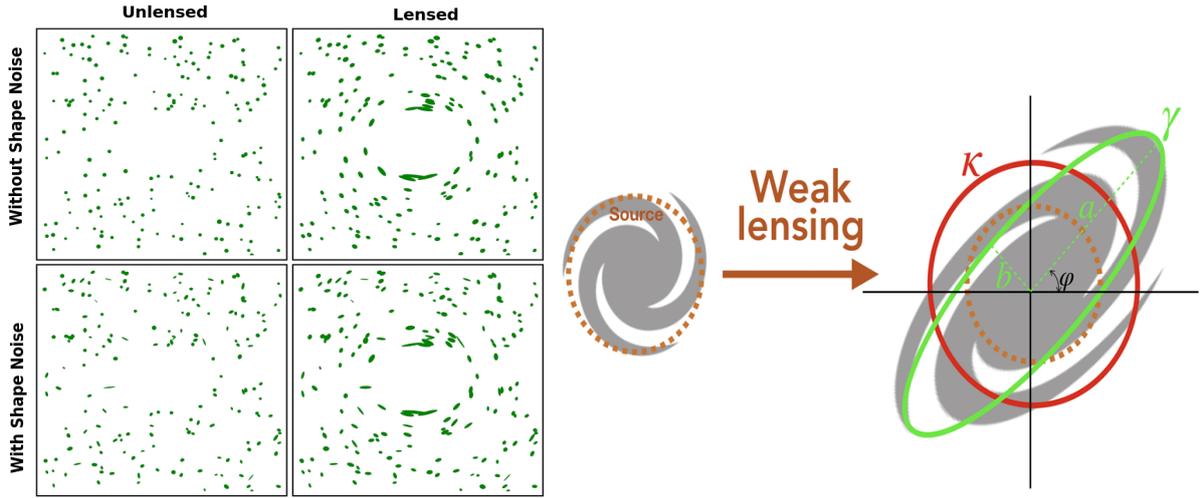

Figure 2.10: *Lensed and unlensed images under the influence of shape noise. Credit: Wikipedia*

Figure 2.11: *Influence of weak lensing for convergence and shear. Credit: Marko Shuntov*

The convergence magnifies the background objects by increasing their size and luminosity, while conserving surface brightness. A mass distribution with a fixed surface-mass density $\sum$ can thus be a more or less efficient gravitational lens, depending on the overall extent of the lens system composed of observer, source, lens and depending on where the lens is located along the line-of-sight. Lensing is most efficient where the critical surface-mass density $\sum_{cr}$ is minimal. In Euclidean space, this would be half-way between the observer and the source. In the curved space-time of the Universe, the location of maximal lensing sensitivity is somewhat closer in redshift to the observer. One of the main features of gravitational lensing is the distortion which it introduces into the shape of the sources. For example, background galaxies can appear as very long arcs in galaxy clusters. The distortion arises because light bundles are deflected differentially. Ideally the shape of the images can be determined by solving the lens equation:

$$\vec{\beta} = \vec{\theta} - \vec{\nabla}\vec{\Psi} \tag{2.33}$$

In particular, if the source is much smaller than the angular size on which the physical properties of the lens change, the relation between source and image positions can locally be linearized. Then, the corresponding angular distance of the image point can be approximated by a first-order Taylor expansion of the lens equation with A as the Jacobian matrix of the linearized lens mapping.

$$\delta\vec{\beta} \approx A\delta\vec{\theta} \tag{2.34}$$

It has the components

$$A_{ij} = \frac{\delta\beta_i}{\delta\theta_j} = \delta_{ij} - \frac{\delta^2\Psi(\vec{x})}{\delta x_i \delta x_j} \qquad with \qquad \frac{\delta^2\Psi(\vec{x})}{\delta x_i \delta x_j} = \Psi_{ij} \tag{2.35}$$

For the physical interpretation of the Jacobi matrix A, it is convenient and instructive to split A into an isotropic and an anisotropic, trace-free part by taking the trace,

$$trA = 2 - \vec{\nabla}^2\vec{\Psi} = 2(1 - \kappa) \tag{2.36}$$

and subtracting it from A by means of the unit matrix I to obtain the shear matrix

$$\Gamma := -\left(A - \frac{1}{2}(trA)I\right)_{ij} = \begin{pmatrix} -\frac{1}{2}(\Psi_{11} - \Psi_{22}) & -\Psi_{12} \\ -\Psi_{21} & \frac{1}{2}(\Psi_{11} - \Psi_{22}) \end{pmatrix} \tag{2.37}$$

with the components

$$\Gamma_{11} =: \gamma_1 = \frac{1}{2}(\Psi_{11} - \Psi_{22}) \quad , \quad \Gamma_{22} = -\gamma_1 \quad , \quad \Gamma_{12} = \Gamma_{21} =: \gamma_2 = \Psi_{12} \tag{2.38}$$

Thus, there exists a so-called shear matrix describing a coordinate rotation where φ is the angle between α and the x-axis. The factor 2 is due to the 2x2 tensor format.

$$\begin{pmatrix} \gamma_1 & \gamma_2 \\ \gamma_2 & -\gamma_1 \end{pmatrix} = \gamma \begin{pmatrix} cos2\varphi & sin2\varphi \\ sin2\varphi & -cos2\varphi \end{pmatrix} \tag{2.39}$$

The shear can be written in complex form:

$$\gamma = \gamma_1 + i\gamma_2 = |\gamma|e^{2i\varphi} = \sqrt{\gamma_1^2 + \gamma_2^2} \tag{2.40}$$

The amplitude describes the amount of distortion, and the phase indicates the distortion direction. We can also define a tangential and a cross component relative to the direction φ:

$$\gamma_t = -Re[\gamma^{-2i\varphi}], \quad \gamma_\times = -Im[\gamma^{-2i\varphi}] \tag{2.41}$$

where the factor 2 again implies that the shear is not a vector but a tensor defined by the trace-free part of the symmetric Jacobian matrix A.
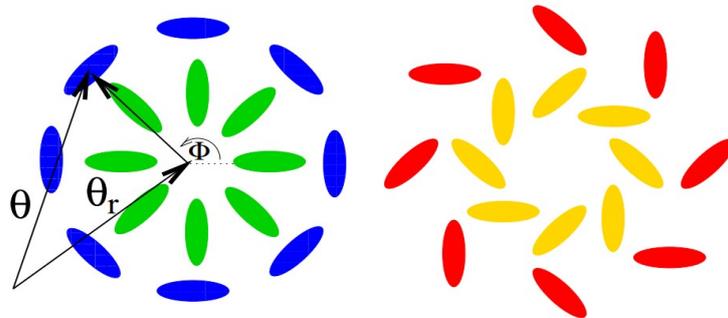

Figure 2.12: *Illustration of the tangential and cross components of the shear, measured with respect to the reference point θ_r at the center of the image. Left: $\gamma_\times = 0$ and $\gamma_t = 1$ (outer ellipses),$\gamma_t = -1$ (inner ellipses). Right: $\gamma_t = 0$ and $\gamma_\times = -1$ (outer ellipses), $\gamma_\times = 1$ (inner ellipses). Credit: [18]*

In this case, in fact, the shear will be always positive and oriented tangentially respect to the center of symmetry, while the cross component will always be null [17].

These manipulations leave the Jacobi matrix in the form

$$A = (1 - \kappa)I - \Gamma = \begin{pmatrix} 1 - \kappa - \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa + \gamma_1 \end{pmatrix} \tag{2.42}$$

The inverse Jacobi matrix determines how sources are mapped on images as linear lens mapping for weak gravitational lensing is invertible

$$A^{-1} = \frac{1}{detA} \begin{pmatrix} 1 - \kappa + \gamma_1 & -\gamma_2 \\ -\gamma_2 & 1 - \kappa - \gamma_1 \end{pmatrix} \tag{2.43}$$

The prefactor in this expression indicates that the solid angle spanned by the image is changed compared to the solid angle covered by the source by the magnification factor

$$\mu = detM = \frac{1}{detA} = \frac{1}{(1 - \kappa)^2 - \gamma^2} \approx 1 + 2\kappa \tag{2.44}$$

Thus, in weak lensing, the magnification of an image is essentially (i.e. to the first Taylor order) determined by the convergence κ, not by the shear γ. The distortion induced by the convergence is isotropic, i.e. the images are only rescaled by a constant factor in all directions. On the other hand, the shear stretches the intrinsic shape of the source along one distinctive direction. For this reason, a circular source, which is small enough compared to the scale of the lens, like that shown in 2.11 is mapped into an ellipse with the properties

$$\epsilon(\vec{\theta}) = \frac{a - b}{a + b} = \frac{2\gamma(\vec{\theta})}{2(1 - \kappa)} \approx \gamma(\vec{\theta}), \quad with \quad a = \frac{r}{1 - \kappa - \gamma}, \quad b = \frac{r}{1 - \kappa + \gamma} \tag{2.45}$$

Galaxy clusters are the largest gravitationally bound structures in the Universe. When assuming an Einasto profile[5] of mass density for spherical stellar systems, the strength of the tangential shearing caused by foreground matter can be calculated by creating a magnification map.
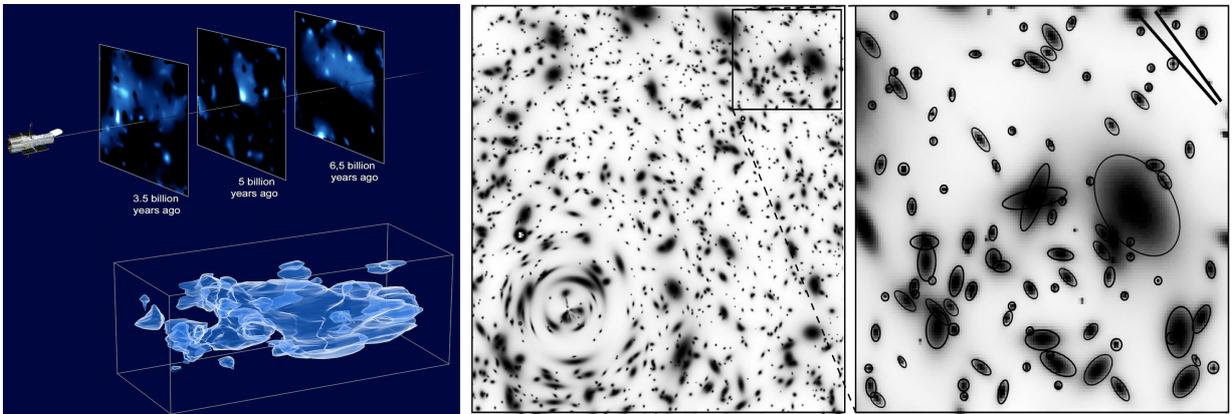

Figure 2.13: *Reconstruction of dark matter distribution in the Universe. Credit: Scienceblogs.com*

Figure 2.14: *Weak lensing recording. Credit: Caltech.edu*

---

[5]https://arxiv.org/pdf/1610.04620.pdf

The fact that galaxy clusters are scattered throughout the universe allows for lensing and magnification of the cosmic shear, as can be seen in figure 2.14. By methods of inversion and ray tracing galaxy-galaxy lensing enables deeper insight ultimately culminating in a view of the Universe many billion years ago. As 80% of a galaxies mass is in the form of dark matter, lensing methods can also be usefull in the exploration and detection of such uncharted accumulations. Abell 1689, CL0024+17, and the Bullet Cluster are among the most prominent examples of lensing clusters. Lensing mass maps can also potentially reveal "dark clusters", containing overdense concentrations of dark matter while providing constraints on models such as MOND. [6]

- **Mircolensing**

  Smaller objects, like individual stars, can also act as gravitational lenses when they pass in front of more distant stars. For a few days or weeks, light from the more distant star temporarily appears brighter because it is magnified by the gravity of the closer object. Distortion itself can not be seen, whereas the amount of light received from a background object changes in time. The effect is small, such that in the case of strong lensing even a galaxy with a mass of more than 100 billion times that of the Sun will produce multiple images separated by only a few arcseconds. Galaxy clusters can produce separations of several arcminutes. In both cases the galaxies and sources are quite distant, many hundreds of megaparsecs away from our Galaxy [19]. Unlike with strong and weak lensing, no single observation can establish that microlensing is occurring. Instead, the rise and fall of the source brightness must be monitored over time using photometry. This function of brightness versus time is known as a light curve. A typical microlensing light curve is shown below:
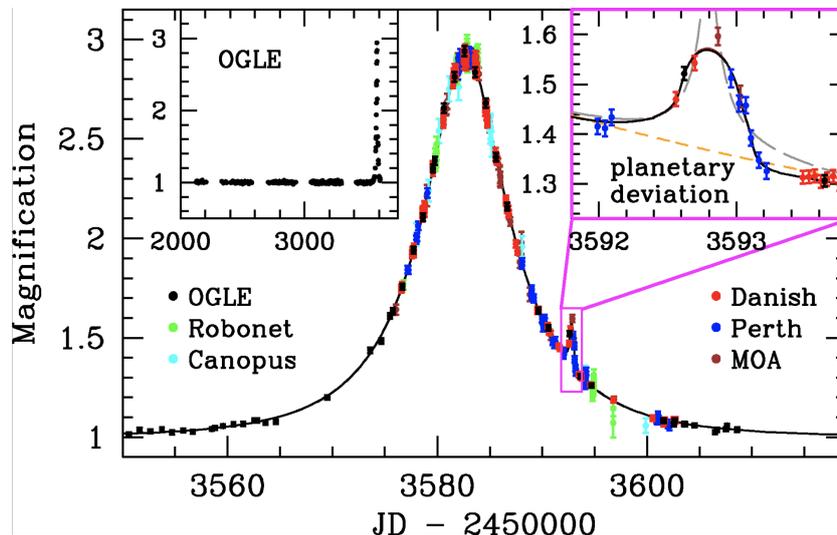

Figure 2.15: *Micro-magnifications measured for stars in orbit around an object of a very large mass. Also known as "Transit method" when observing Exoplanets. Credit: OGLE*

Thus mapping the Universe in higher event definition, modified microlensing also serves as a efficient and succesfull method when searching for exoplanets.

---

[6]Modified Newtonian Mechanics

Chapter 3

Cosmological properties

The physical universe encompasses all of spacetime and its contents. Such contents are defined as all of the observable energy. This includes planets, stars, intergalactic dust and even electromagnetic radiation.

Until about thirty years ago, astronomers thought that the universe was composed almost entirely of ordinary atoms. However, in the past few decades, there has been ever more evidence accumulating that suggests there is some new form of matter in the universe which we can not see. This fact calls into question the previous understanding of the universe and accordingly requires a comprehensive examination of the individual categories and presentation of their relative distribution among one another, taking into account the latest discoveries. For this reason, the elementary components of the known universe, as well as newly acquired hypotheses will be presented in this chapter.

3.1   Baryonic matter

For the simple reason that most of the mass of an atom is concentrated in its nucleus, which is made up of baryons, astronomers often use the term baryonic matter to describe so-called ordinary matter. This ordinary matter includes all accumulations of mass visible to mankind, meaning the definition is based on electromagnetic interactions. As a matter of fact, the great majority of ordinary matter in the universe is unseen, since visible stars and gas inside galaxies and clusters account for less than 10 % of the ordinary matter contribution to the mass-energy density of the universe. Ordinary matter commonly exists in four states: solid, liquid, gas, and plasma and is composed of two types of elementary particles: quarks and leptons. Protons and neutrons are both made up of quarks, where as electrons are categorised among the lightweight leptons [20] [21].

In astronomical length scales, gravity is the dominant fundamental interaction as its effects are cumulative, unlike electromagnetism (EM), where positive and negative charges cancel each other out. As the two remaining interactions, the weak and the strong nuclear forces decline very rapidly with distance, their effects are confined mainly to sub-atomic length scales. As baryons consist of quarks, they succumb to all of the forces mentioned above, excluding EM if not charged.

All known particles are officially categorized into groups of the same attributes using the so-called standard model of particle physics.
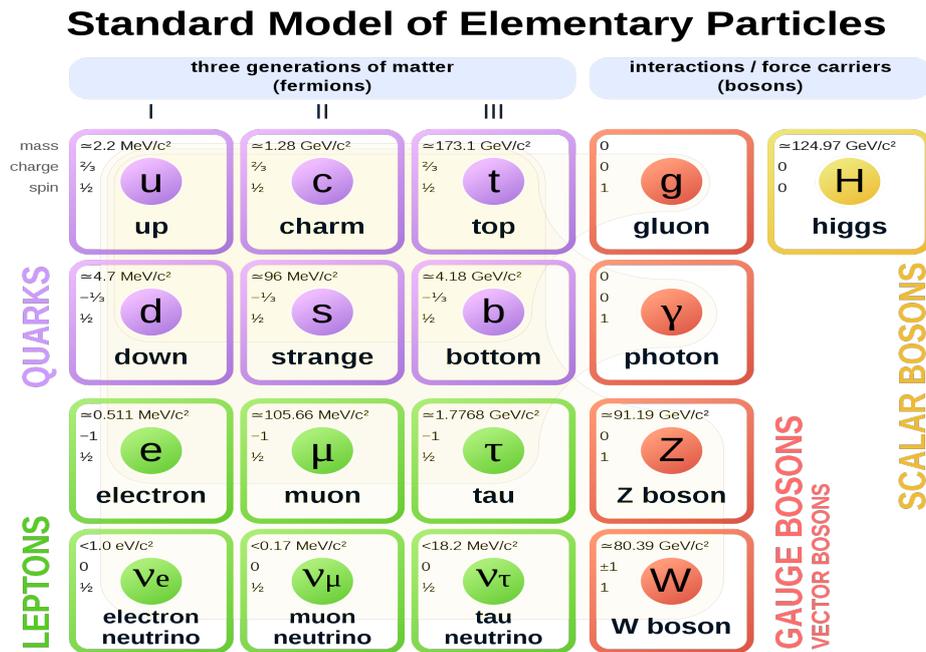
Standard Model of Elementary Particles

	three generations of matter (fermions)			interactions / force carriers (bosons)	
	I	II	III		

| | ≈2.2 MeV/c² ⅔ ½ **u** up | ≈1.28 GeV/c² ⅔ ½ **c** charm | ≈173.1 GeV/c² ⅔ ½ **t** top | 0 0 1 **g** gluon | ≈124.97 GeV/c² 0 0 **H** higgs |

mass
charge
spin

QUARKS:
≈4.7 MeV/c² −⅓ ½ **d** down | ≈96 MeV/c² −⅓ ½ **s** strange | ≈4.18 GeV/c² −⅓ ½ **b** bottom | 0 0 1 **γ** photon

LEPTONS:
≈0.511 MeV/c² −1 ½ **e** electron | ≈105.66 MeV/c² −1 ½ **μ** muon | ≈1.7768 GeV/c² −1 ½ **τ** tau | ≈91.19 GeV/c² 0 1 **Z** Z boson

<1.0 eV/c² 0 ½ **νe** electron neutrino | <0.17 MeV/c² 0 ½ **νμ** muon neutrino | <18.2 MeV/c² 0 ½ **ντ** tau neutrino | ≈80.39 GeV/c² ±1 1 **W** W boson

GAUGE BOSONS / VECTOR BOSONS

SCALAR BOSONS

Figure 3.1: *Standard Model of Particle Physics. Credit: Wikipedia*

This model lists all so-far experimentally confirmed existing particles which compose matter. These are grouped in fermions and bosons, based on their spin. Furthermore fermions can be split into two collections, quarks and leptons, according to their electromagnetic charge. Within each collection the particles are ranked from left to right in mass. Corresponding "antimatter" partners as well as the force particles which mediate interactions by energy transferring are also listed. Such mediation particles, in particular the mass they carry, become significant when comparing the sum of a quark arrangement with the mass of the corresponding nucleus. Mediation particles are photons (γ) for electromagnetic interactions, gluon (g) for strong interaction and the gauge bosons Z^0, W^-, W^+ for weak interaction. However, the cause of the existence of the interactions could be traced back to the ambivalent symmetries. Due to this, particles of matter maintain a consistent appearance at all times and in all places.

The standard model does not, however, accommodate gravity. A true force particle has not yet been attained. It can therefore not be a complete description of particle physics. As previously mentioned, other recent discoveries regarding phenomena such as dark matter indications demand for the model to be updated. Suitable candidates for particles that make up dark matter are supersymmetric particles.

Overall the universe appears to have much more matter than antimatter. This occurrence, also known as baryon asymmetry, is one of the most important phenomena of particle physics not yet understood, since it cannot be explained by the standard model of elementary physics. This imbalance between matter and antimatter is partially responsible for the existence of all matter existing today, since matter and antimatter, if equally produced at the Big Bang, would have completely annihilated each other and left only photons as a result of their interaction.

3.2 Dark Matter

Based on the studies of the observable galaxies, the discovery has been made that the universe contains much more matter than initially thought to be accounted for by the previously explained ordinary objects. A wide range of strong indirect evidence exists, which suggests that additional material, not explained within the frame of the standard model, lies scattered all over the universe. Today it is well established that dark matter makes up about 26% of the energy density of the Universe, while also being about six times more abundant than ordinary matter [22]. The fundamental nature of such hypothetical matter, however, remains one of the greatest mysteries in modern astrophysics, as it neither emits nor absorbs light and therefore can not be detected directly. Unveiling which particle accounts for the majority of the matter in the Universe is an open question at the interface of particle physics and cosmology. Reasoning to defend the hypothesis may be derived from its gravitational effects on visible matter and all encompassing structures throughout the Universe [23]. Some of the experimental facts that support the idea of dark matter are shown in fig3.2 and will be briefly discussed in the following.
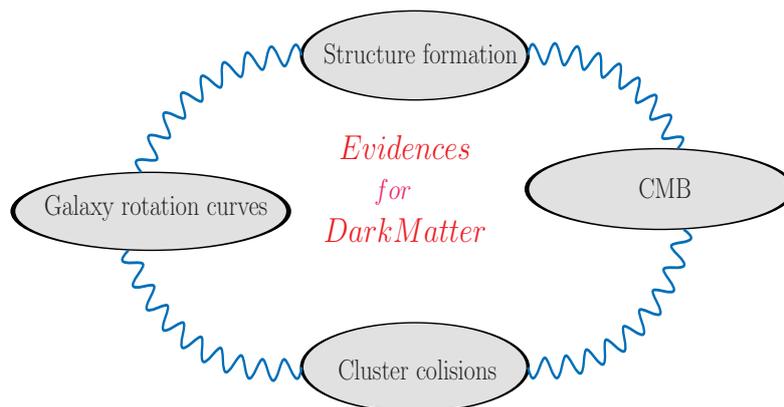
Figure 3.2: *Some of the most important evidences for Dark Matter.*

3.2.1 Galaxy rotation curves

Throughout space, astrophysical objects of all sizes swirl and orbit. To keep these objects tightly bound together, the gravitational pull felt by an object must be strong enough to balance the energy it has due to its motion. The rotation curve of a Galaxy (cluster) is the profile of the circular velocity of the stars (galaxies) around the mass center of the system. Such profile allows to calculate the mass distribution of the galaxy (cluster) at hand. Historically, the relation between the mass distribution and the rotation curve was first proposed by Fritz Zwicky in 1933. He analyzed the velocity dispersion of the galaxies in the Coma cluster, assuming that the outer galaxies were in circular motion around its mass center. He applied the virial theorem[1] to the Coma cluster, in order to estimate its mass and found that roughly 800 galaxies should exhibit velocities of 80 km/h, however, the observed velocity dispersion was approximately 1000 km/h [24].

---

[1]an equation which relates the average kinetic energy of a system to its total potential energy

He found that the mass from the luminous matter was not enough to keep the cluster bound, and was several times smaller than the inferred gravitational mass.
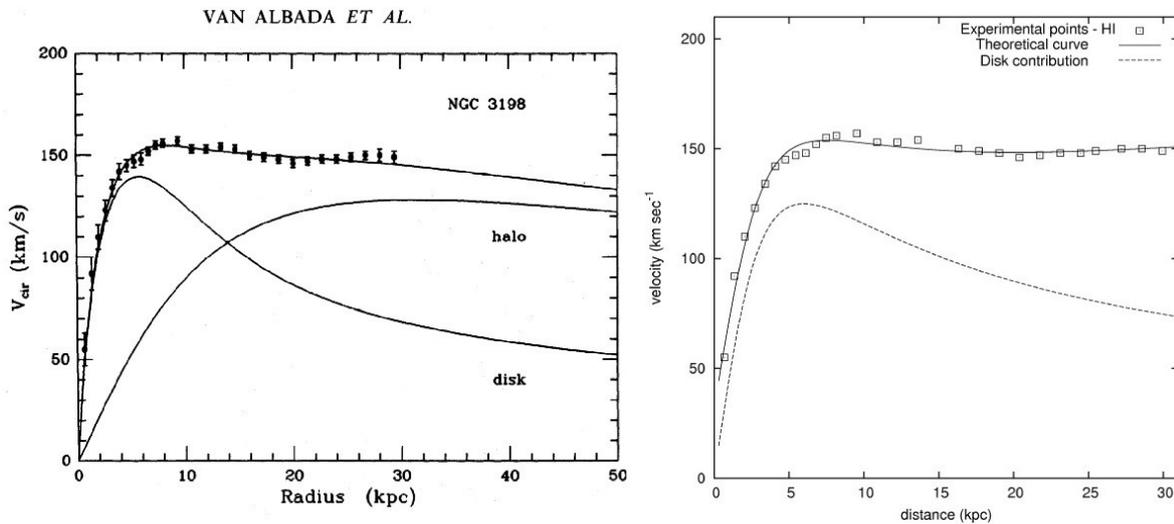

Figure 3.3: *Observed velocities versus distance from the center of galaxy NGC 3198. The theoretical prediction before observations followed the trend labeled "disk", but observations (black squares) showed constant, rather than decreasing velocity. Adding a contribution from a dark matter halo (center line) makes the theory match predictions. Credit: Van Albada et al. (L), A. Carati, via arXiv:1111.5793 (R).*

The problem was known as the Galaxy rotation problem. The general idea under this problem is that when Newton mechanics are used to explain the velocity distribution of the stars and visible gas in a Galaxy, the obtained profile (disk) does not match with the observed behaviour which is measured with applied astrophysical techniques such as the mass-to-light ratio[2] and the distribution of stars in the spiral galaxies. This problem is solved, when the existence of dark hidden mass is assumed. If instead a large fraction of the galaxy's mass resided in a diffused dark matter 'halo' which extended well beyond the edges of the luminous matter, the observed galactic rotation curves could be explained. Dark hidden mass is present in the galaxy with a special distribution which governs its gravitational behaviour.

3.2.2 The Cosmic Microwave Background (CMB)

The CMB is the oldest snapshot of the Universe. It corresponds to the thermal radiation of the Universe approximately 380.000 years after the Big Bang ($z \approx 1100$, $T \approx 3000$ K). This radiation was generated in a time in the thermal history of the Universe called recombination or "time of the last scattering", which was the time when the electrons and protons formed bound states and created the neutral hydrogen in the Universe. This discovery, made by the American radio astronomers Arno Penzias and Robert Wilson in 1964, was considered a test of the Big Bang theory and the cosmological lambda cold dark matter model (Lambda-CDM or Λ-CDM).

---

[2]The mass-to-light ratio (Υ) is the relation between the total mass of a galaxy and its luminosity. In astrophysics the reference value is the mass-to-light ratio of the sun, for that reason for big objects dominated by DM have a big mass-to-light ratio.
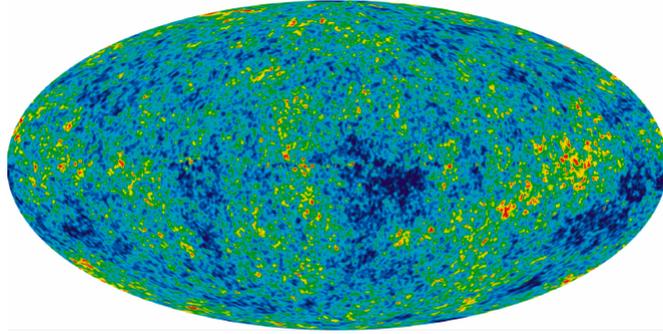
Figure 3.4: *All-sky picture of the infant Universe created from nine years of WMAP data. The image reveals 13.77 billion-year-old temperature fluctuations. This image shows a temperature range of ±200 microKelvin. Credit: NASA / WMAP Science Team WMAP #121238 Image Caption 9 year WMAP image of background cosmic radiation (2012).*

In general, the CMB map has a thermal black body spectrum at a temperature of 2.72548 ± 0.00057 K with a spectral radiance of 160.23 GHz, i.e. in the microwave range of frequencies. Even more, this spectrum shows tiny temperature fluctuations, which correspond to regions of slightly different densities which were the seeds of all the structures as the galaxies which are present in the Universe. The existence of dark matter leaves a characteristic imprint on CMB observations, as it clumps into dense regions and contributes to the gravitational collapse of matter, but is unaffected by the pressure from photons. This oscillating effect in the CMB can be presented in the form of a power spectrum [25].
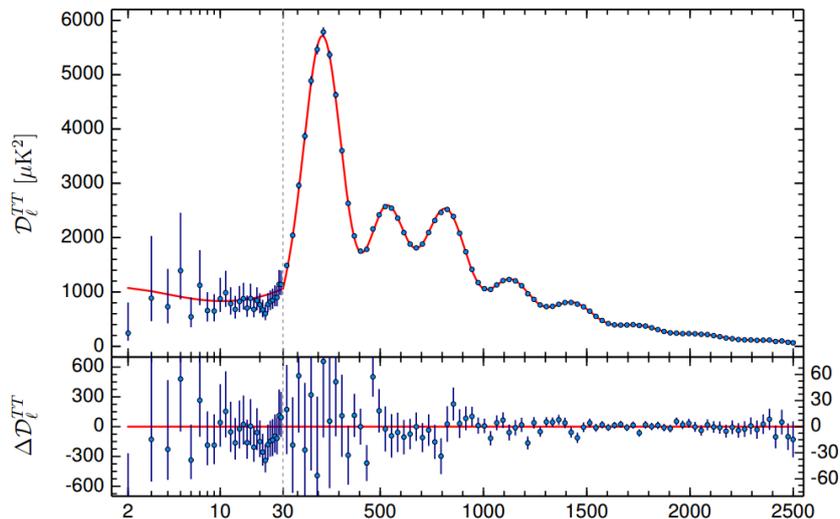
Figure 3.5: *Planck 2015 temperature power spectrum [22]. For multipoles $l \geq 30$ is show the maximum likelihood frequency-averaged temperature spectrum. The best-fit, i.e. the ΛCDM theoretical spectrum is fitted in the upper panel. Residuals with respect to this model are shown in the lower panel. The error bars show $\pm\sigma$ uncertainties.*

A careful analysis of the spectrum above shows some special features. The first peak visible is also known as the acoustic peak and its angular scale $l \approx 200$(1°in galactic coordinates) determines the curvature of the Universe. For instance, the latest Planck data combined with

gravitational lensing and baryon acoustic oscillation (BAO) indicate that k = 0.000 ± 0.005 with 95% confidence level. This means that the Universe is spatially flat at high precision. The ratio between the second and the first peak determines the baryon density. Finally, the third peak in combination with the first and the second peak can be used to obtain information about the dark matter density in the Universe. The Wilkinson Microwave Anisotropy Probe (WMAP) was the first instrument to measure the CMB power spectrum through the first peak of oscillations, and showed that the existence of dark matter is favored.

3.2.3 Bullet Cluster

In 2006 a group of astronomers studied the merging of two clusters of galaxies 1E 0657-558 collectively known as the bullet cluster [26]. The collision which created this formation is estimated to have taken place ∼ 100 million years ago. In general, they found that there are two concentrations of galaxies separated by ∼ 0.72 Mpc. The right galaxy is continuously moving away at speeds of ∼ 4700 km s^{-1} and thus creating the name giving bow shock. This observation depicts a single system in which the baryonic matter was separated from the mass center of each cluster involved in the collision.
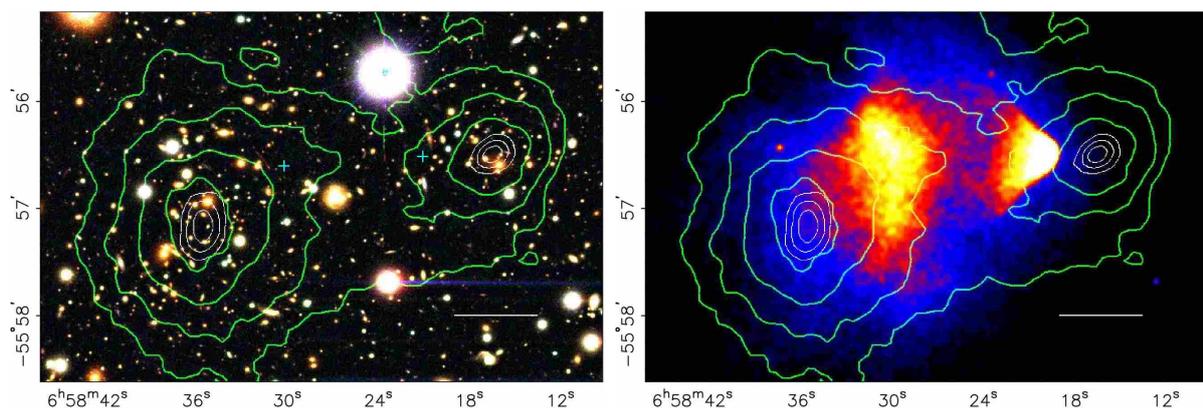
Figure 3.6: *The bullet cluster. The green contours are the reconstruction with gravitational lensing which is proportional to the mass of the system. The white bar represents a distance of 200 kpc at the location of the cluster. The colored map on the right shows the same image seen in X-ray for the merging cluster. It was taken with the Chandra satellite after 500 seconds of exposure. Credit: [26].*

By using methods of gravitational lensing, the astronomers were able to reconstruct the contours of the mass projected for the system. They are represented by the green contours in fig. 3.6. Additionally Chandra observations of the emitted X-rays by the baryonic plasma of the system are illustrated as blue crosses in fig. 3.6. An extremely hot gas of particles pervades the space between each galaxy in a cluster, which accounts for about 90% of the mass from ordinary matter (rather than stars). When two galaxy clusters collide, the gas particles become even hotter from crashing into each other, causing an increase in brightness of the X-ray emission. From this we can tell how energetic the gas is and where it is located. Yellow and red parts of the image represent the baryonic plasma, which emits the X-rays. Clearly the plasma distribution does not match the mass distribution found with the gravitational lensing

reconstruction. If the clusters were entirely comprised of ordinary matter, the location of mass from the optical observations and the location calculated from gravitational lensing in the bullet cluster should overlap. Instead, the observations showed a glaring inconsistency. As a conclusion, during the galaxy merger, the majority of mass in the galaxies behaves almost collisionless, whilst decoupling from all the baryonic matter. Similar to two clouds colliding and passing through each other with only the baryonic articles colliding. According to this interpretation, the principal component of the mass of the system does not interact. It is dark, not baryonic, and corresponds to the green contours shown in fig. 3.6. Note that this observation is in favor of the DM interpretation as a particle. Even more, an alternative explanation using theories of MOND does not predict an offset between mass and light and could fail to explain this observation of the bullet cluster [27].

3.3   The search for Dark Matter

- **Direct Detection**

  The idea of direct detection of DM is based on the fact that DM particles, so-called non baryonic Weakly Interacting Massive Particles (WIMPs) or Gravitationally Interacting Massive Particles (GIMPs) are capable of collision with nucleons. Many experiments to directly detect and study dark matter particles are being actively undertaken, but none have yet succeeded. Until now, amongst the experiments for direct detection of DM, the most restrictive is the Large Underground Xenon experiment (LUX) 2 , which is located 1,510 m underground at the Sanford Underground Laboratory, South Dakota. It is operated underground to reduce the noise signal caused by high-energy cosmic rays at the Earth's surface.
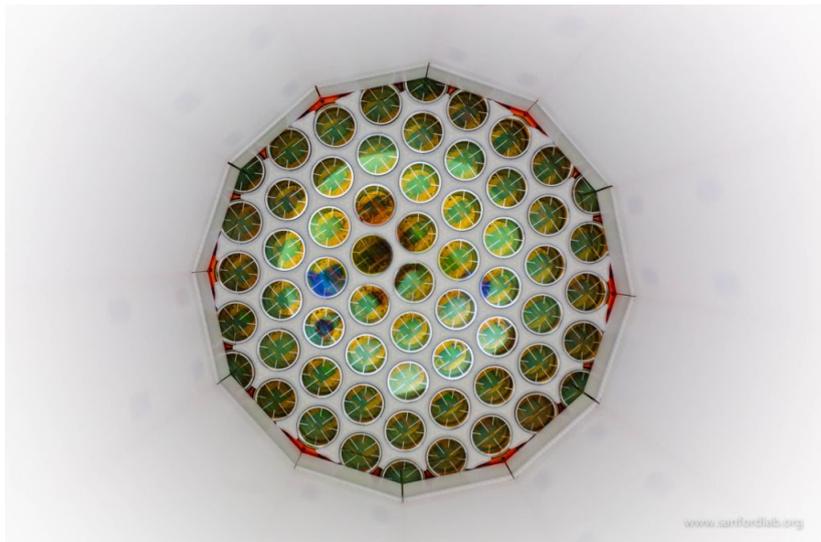
Figure 3.7: *The view from inside the Large Underground Xenon (LUX) dark matter detector, which is nearly a mile underground below the Black Hills of South Dakota. The upgraded detector just finished its 20-month run without finding dark matter activity. Credit: Matthew Kapust. Copyright © South Dakota Science and Technology Authority*

It is hoped that the interaction between DM and the liquid Xenon in the detector generates 175 nm ultraviolet photons and some electrons.

This would be a first step towards confirming the basis of the theoretical assumptions regarding the all encompassing presence of dark matter. Many supersymmetric models offer dark matter candidates in the form of the WIMPs Lightest Supersymmetric Particle (LSP). Separately, heavy sterile neutrinos, a slower form of neutrino that does not interact through the weak force, exist in non-supersymmetric extensions to the standard model, which explain the small neutrino mass through the seesaw mechanism.

- **Indirect Detection**

  Indirect detection efforts typically focus on locations where WIMP dark matter is thought to accumulate the most: in the centers of galaxies and galaxy clusters, as well as in the smaller satellite galaxies of the Milky Way. These are particularly useful since they tend to contain very little baryonic matter, reducing the expected background from standard astrophysical processes. Typical indirect searches look for excess gamma rays, which are predicted both as final-state products of annihilation, or are produced as charged particles which interact with ambient radiation, via inverse Compton scattering. Experiments which have placed bounds on WIMP annihilation, via the non-observation of an annihilation signal, include the Fermi-LAT gamma ray telescope and the VERITAS ground-based gamma ray observatory. Although the annihilation of WIMPs into standard model particles, also predicts the production of high-energy neutrinos, their interaction rate is too low to reliably detect a dark matter signal at present.

Chapter 4

Data Processing

Throughout this chapter methods used for data analysis will systematically be introduced, giving insight into how deep learning techniques will be applied later in this thesis. The methodology of computer vision based on multilayer networks will be illustrated in the example of the open-source software libraries Keras and TensorFlow.

4.1   Artificial Neural Network

It might appear rather surprising that Artificial Neural Networks (ANN's) were in fact first introduced in 1943 by the neurophysiologist Warren McCulloch and mathematician Walter Pitts. Their ideas were based on inspiration they had taken from the Biological Neural Network (BNN) within mammals. Research on the cerebral cortex of the brain has revealed, that nerve cells appear to be organised in consecutive layers, creating an architecture capable of processing information. These nerve cells, also known as neurons, are composed of a cell body containing the nucleus and most of the cell's complex components, many branching extensions called dendrites, plus one very long extension called the axon. Near its extremety the axon splits of into many branches called telodendria which then culminate in synaptic terminals, which are connected to the dendrites of further neurons. Neurons generate electrical impulses called action potential which then travel along the axon resulting in the synapses releasing chemical substances called neurotransmitters, which can be either inhibitory or stimulating [28] [29].
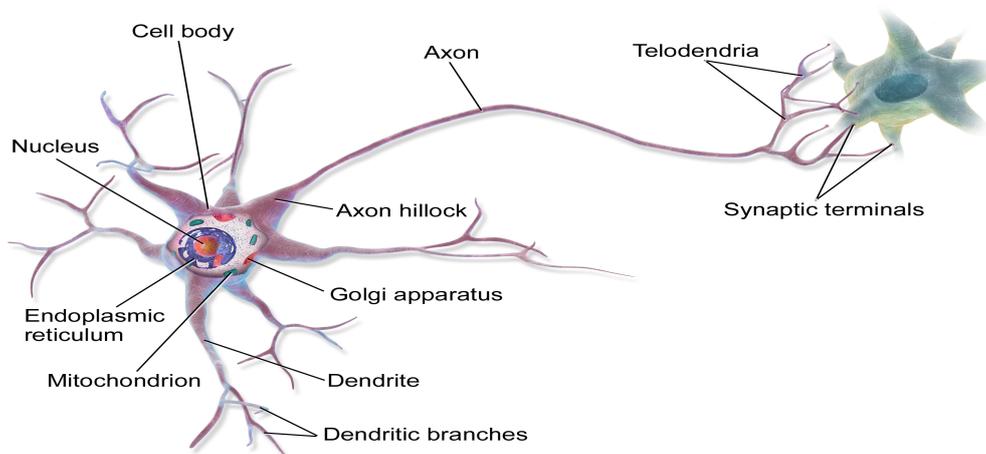
Figure 4.1: *Functionality and set up of a biological neuron. Credit: Wikipedia*

This results in the change of the electrical potential of connected neurons. In other words, the propagation of information within the cell is carried out by the action potentials, whereas the transmission to further cells is triggered when a certain threshold is overcome. Today it is assumed that there are approximately 86 billion neurons inside the brain, with each of them being connected to 1000-10000 others [30].

4.2 Single & Multilayer Perceptron

The workflow of BNN's was first artificially modelled by Franck Rosenblatt's Threshold Logic Units (TLU's) or simply perceptrons. These replace the synaptical signal with weighted inputs to detect patterns and a neuron specific bias, which establishes when the neuron becomes meaningfully active. It also replaces in-cell activation potentials with step functions and the axon with an output. The TLU then functions as a multioutput classifier. The full potential of ANN's has only recently been discovered, as more complex architectures rely on increasing computing power. This is partially due to Moore's law[1].
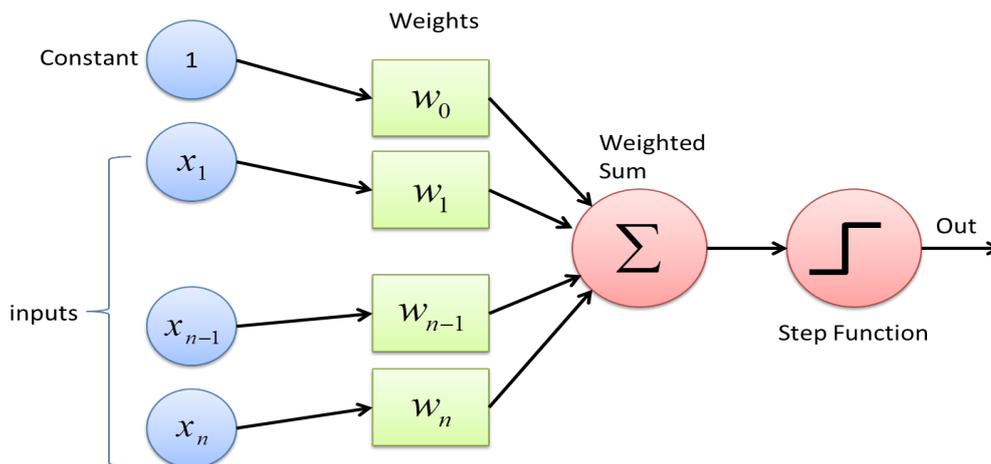
Figure 4.2: *Structure and mode of operation of an artificial Perceptron. Credit: Towards Data Science*

Unfortunately Perceptrons also have some serious weaknesses, such as the fact that they are incapable of solving trivial classification problems, also known as *exclusive OR or XOR*. Some of these limitations, including the XOR-problem, can however be eliminated by stacking multiple Perceptrons. The resulting ANN is called a Multilayer Perceptron (MLP). An MLP is usually composed of:

- one input layer - every TLU represents a feature for example the pixels of an image to be classified (e.g. a black and white image with pixel values ranging from 0-255)

- one or more *hidden layers* of TLUs, which perform filter operations on pixel level

- one output layer which represents the target value depending on the application area. In case of classification the number of output nodes will equal the number of classes

---

[1]http://www.mooreslaw.org

In general a distinction is made between two types of neural networks [20]:

I. **Feedforward neural networks (FFNN)**

FFNNs are characterized by unidirectional connections, from input to output. Put simply, one goes from one layer to the next as there are no connections back to a previous layer. These layers are also known as *fully connected* or *dense* layers. During training they learn by adapting their connection weights by calculating the difference of its own response with the supposed outcome of the training set. The largest area of application is found with Convolutional Neural Networks (CNNs), which are increasingly used in image recognition.

II. **Feedback neural networks (FBNN)**

Within feedback network the neurons have a direct connection to their immediate predecessors. They are mainly used for processing problems in which temporal aspects have to be taken into account explicitly. Therefore these systems are of no further interest for this thesis.

4.3 Deep Learning

One of the most common AI techniques used for processing big data is machine learning, a self-adaptive algorithm that gets increasingly better at analysis of patterns with experience or with newly added data. Deep learning is a subset of machine learning in AI that imitates the workings of the human brain in processing data and creating patterns for use in decision making. While traditional programmes build analysis with data in a linear way, the hierarchical function of deep learning systems enables machines to process data with a non-linear approach [31].
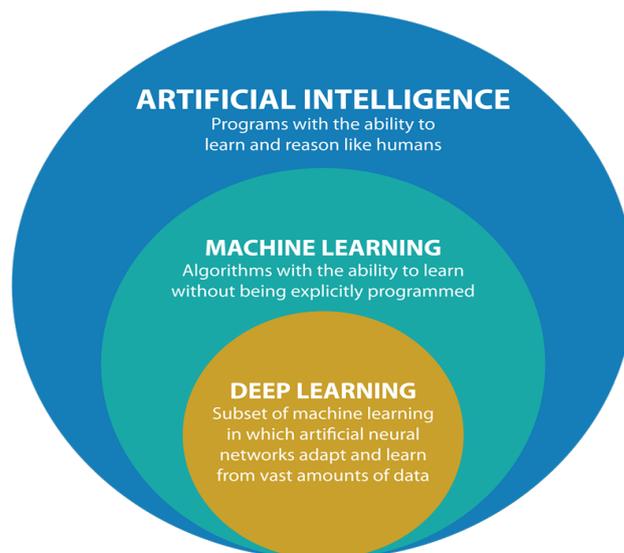
Figure 4.3: *Division of the relationship of AI, ML and Deep learning. Credit: CLEANPNG*

As explained previously each level of an MLP architecture learns to transform its input data into a slightly more abstract and composite representation. The closer the level is to

the output layer, the higher the complexity of features, which can be recognized, becomes. The network moves through the layers calculating the probability of each output. If the network did not accurately recognize a particular pattern, an algorithm adjusts the weights of the connections as well as the biases. That way the algorithm can make certain parameters more influential, until it determines the correct mathematical manipulation to fully process the data and converge. Each mathematical manipulation as such is considered a layer, and complex networks have many layers, hence the name "deep" networks.

4.3.1 Learning techniques

To optimize the neuronal network until it converges with the correct results, it has to be trained. Any implementation of neural networks includes parameters that have to be selected before training. Depending on the implementation, these can resemble the number of neurons, number of layers, number of epochs, batch size and learning rate. If parameters are not chosen correctly, results might either be inaccurate, over-fitted, under-fitted or show unexpected convergence behaviour, such as when "stuck" in a local minima [28].
Depending on the training of neural networks two main types have to be distinguished:

- supervised learning from annotated data

- unsupervised learning - fully automatic learning from non-annotated data

4.3.2 Backpropagation

In order to be able to improve all levels of a MLP, the so-called *loss function* is used.
It should be noted that the selection of the loss function is based on the area of network application. Examplamentary the *Mean Squared Error* (MSE) function is defined as:

$$E(\mathbf{y}) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \geq 0 \tag{4.1}$$

Here E is the error, n is the number of training examples and y and $\hat{y}$ represent the desired and calculated target values. In order to learn in a further iteration step, the network tries to minimize the discrepancy between setpoint and target value through backpropagating the error. The weight adjustment is calculated using the stochastic gradient descent method[2]:

$$w_{i_{new}} = w_{i_{old}} - \eta \cdot \frac{\delta E}{\delta w_i} \tag{4.2}$$

The new weights depend on the error dimension as well as its partial derivative. Latter is also multiplied by the learning rate which determines the speed of adjustment. These steps are repeated until the network converges in the solution. In order for this algorythm to work properly the classical step function needs to be replaced with a well-defined non zero derivative function such as the logistic (sigmoid) function or the hyperbolic tangent function.

---

[2]After completing a batch in training the negative gradient of all variables is calculated and summed for every output node. The resulting tweak to each variable minimizes the error for the next training batch.

4.4 Computer vision

One of the most powerful and compelling interdisciplinary fields of artificial intelligence is computer vision. It focuses on training computers to interpret and understand visual images. Using digital images from cameras, videos and deep learning models, machines can accurately identify and classify objects — and then react to what they "see". Based on skills such as acquiring and processing it will analyse data sets, check for shapes, borders and structures and numerical information or simply output form of decision. It involves the development of a theoretical and algorithmic basis to achieve automatic visual understanding.
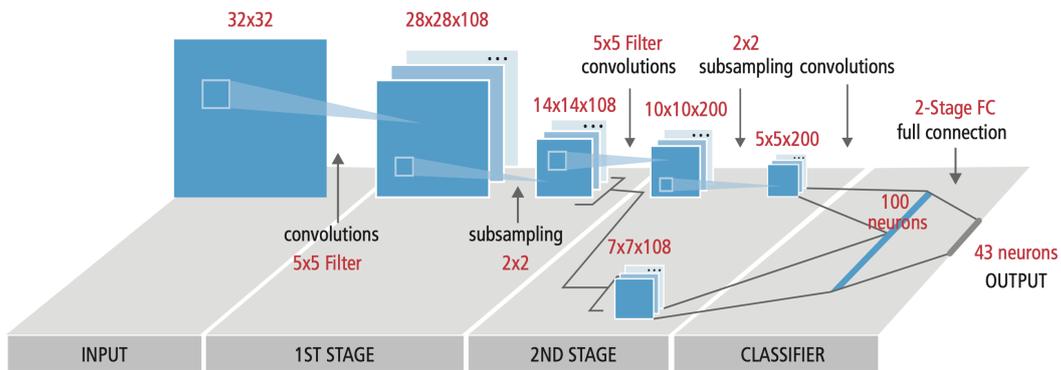
Figure 4.4: *Multilayer setup with Dense layer before output. Credit: [32]*

Usually optical pattern recognition can be achieved with fully connected layers. However, images with more pixels, i.e. more information, overstrain the computing capacity as millions of connections would be necessary. This is why so-called convolutional layers are used in visual computing. Neurons of such a convolutional layer are not connected to every pixel of the input layer but only to pixels in their receptive field. In turn, each neuron in a second layer is only connected to neurons within a certain area of the previous layer. This way the network is able to concentrate on small low-level features in lower layers, followed by high-level features in latter layers.
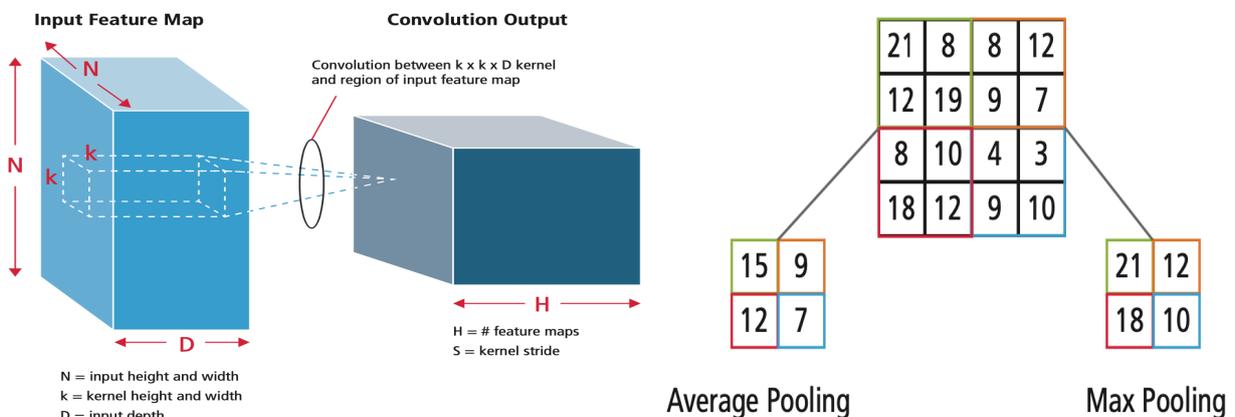
Figure 4.5: *Convolutional layer including pooling filters extract features from input. Credit: [32]*

Figure 4.6: *Valid padding with a 2x2-filter. Credit: [32]*

This way a neuron located in row i, column j of a given layer is connected to the outputs of the neurons in the previous layer located in rows i to $i + f_h$ - 1, columns j to $j + f_w$ - 1, where f_h and f_w are the height and width of the receptive field (see fig. 4.6). It is also possible to connect a large input layer to a much smaller layer by spacing out the receptive fields. This dramatically reduces the model's computing complexity. By representing a neurons weight by an image the size of the receptive field, filters can be applied in order to enhance the classification process. Thus, a layer of neurons using the same filter will output a feature map f_k, highlighting the areas of an image which activate the filter the most. This way the output of a neuron in a convolutional layer, with preceding neurons of weight x, can be calculated as:

$$z_{i,j,k} = b_k + \sum_{u=0}^{f_h} -1 \sum_{v=0}^{f_w} -1 \sum_{n=0}^{f_k} -1 \cdot x_{i,j,k} \cdot w_{i,j,k} \tag{4.3}$$

4.4.1 Tensor Flow

TensorFlow is a powerful library for numerical computation. It was developed by the Google Brain team and is the most popular Deep Learning library to date. At the lowest level Tensorflow operations such as tf.data, tf.io or tf.image for data loading and processing are implemented in C++. Many of these operations have implementations called kernels. Each kernel is dedicated to a specific device type, such as CPUs, GPUs and even TPUs (tenser processing unit). This is particulary helpful as GPUs and TPUs speed up computations dramatically by running calculations parallel across multiple threads.
The next step in the hierarchy is Tensorflows DEE on which all low-level APIs (Application Programming Interface) are run. These include graphs and tensors which can be implemented in programming languages such as python. More advanced mid-level API's may include convolutional layers or functions like the Mean Absolute Error. High-level API's allow to self design neural networks as it connects mid and low-level APIs with simplified functions. Amongst other things the high-level APIs include ready-made estimators, which can already be used for certain tasks (e.g. classification or linear regression) without major adjustments.

4.4.2 Keras

As already described, TensorFlow is a powerful framework, but due to its complexity it is not always user-friendly. The founder of Keras, François Chollet, had the idea of implementing a generically independent framework with several backends[3] as part of the ONEIROS (Open-ended Neuro-Electronic Intelligent Robot Operation System) project.
In 2017 Keras was officially added as one of TensorFlows Deep Learning high-level APIs. Tf.keras[4] allows users to easily build, train, evaluate and execute all sorts of neural networks, while focusing on being modular and extensible. It contains numerous implementations of commonly used neural network building blocks such as layers, objectives, activation functions, optimizers, and a host of tools to simplify working with image data.
The development of the ANN will be implemented via the high-level TensorFlow API tf.keras.

---

[3]Microsoft Cognitive Toolkit (CNTK), Theano, Tensorflow
[4]https://keras.io/

Chapter 5

Cluster map modeling

The appearance of night-sky phenomena due to gravity lensing can often be linked to massive objects by methods of raytracing or N-body simulations. Besides the source being of baryonic nature, dark matter distributions can also be traced. Based on gravitational interaction statistical weak lensing serves as precision measurement for dark matter detection. With the necessity for complex numerical calculations regarding the analysis of the cosmic shear, an alternative approach concerning artificial intelligence has recently come into focus. In order for a neural network to recognize patterns and detect mass distributions, it has to be trained. This is best accomplished by generating simulated test data which will be introduced in this chapter.

5.1 Dark matter density profiles

The hypothesis for CDM structure formation begins with density perturbations in the Universe which grow linearly until they reach a critical density, after which they would stop expanding and collapse to form gravitationally bound dark matter halos. These halos would continue to grow in mass, either through accretion of material from their immediate neighborhood, or by merging with other halos. This cosmological structure is best approximated by a pseudo-isothermal model. However, it cannot be a complete description, as the enclosed mass fails to converge to a finite value as the radius tends to infinity. Numerical simulations of the equilibrium configuration of dark matter halos produced in simulations of collisionless dark matter particles, have resulted in the following universal density profiles [33].

5.1.1 Einasto profile

Originally proposed to describe stellar components of galaxies in 1965, the Einasto profile is characterised by a double power-law logarithmic slope:

$$\gamma(r) = -\frac{dln\rho}{dlnr}(r) \tag{5.1}$$

which, when integrated, leads to the general density profile

$$\rho(r) = \rho_{\mathrm{s}} exp\left(-\frac{2}{\alpha}\left[(\frac{r}{R_{\mathrm{s}}})^{\alpha} - 1 \right] \right) \tag{5.2}$$

where α is known as *Einasto index*, R_{s} is the scale radius and ρ_{s} stands for the halo-density at R_{s}. Based on the Einasto profile, further double power-law density profiles, such as the NFW profile, which indicated the existence of a universal density profile for dark matter halos resulting from the generic dissipationless collapse of density fluctuations emerged [34].
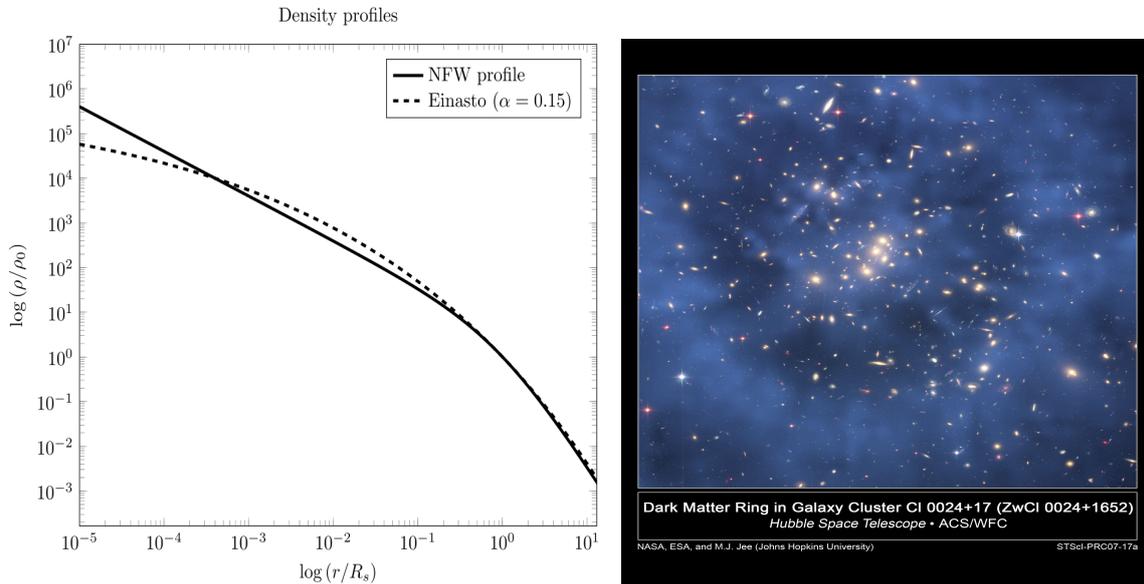
Figure 5.1: *NFW plotted against an Einasto profile. Credit: Wikipedia*

Figure 5.2: *Abell cluster picture ESA for detecting dark matter with assumption of NFW halo density. Credit: NASA*

5.1.2 Navarro-Frenk-White profile

The Navarro-Frenk-White (NFW) profile describes a spatial mass distribution of dark matter halos, fitted to N-body simulations. The density is given as a function of radius:

$$\rho(r) = \rho_{\mathrm{crit}} \frac{\delta_{\mathrm{c}}}{\frac{r}{R_{\mathrm{s}}}\left(1 + \frac{r}{R_{\mathrm{s}}}\right)^2} \tag{5.3}$$

where $\rho_{\mathrm{crit}} = \frac{3H(z)^2}{8G\pi}$ is the critical density at the redshift z and $H(z)$ is Hubble's parameter at that same redshift. The scale radius $R_{\mathrm{s}} = \frac{r_{200}}{c}$ is a characteristic radius of the cluster, c is a dimensionless number known as the concentration parameter, and

$$\delta_{\mathrm{c}} = \frac{200}{3} \frac{c^3}{ln(1+c) - \frac{c}{(1+c)}} \tag{5.4}$$

is a characteristic overdensity for the halo. The virial radius, r_{200}, is defined as the radius inside which the mass density of the halo is equal to $200\rho_{\mathrm{c}}$ (see, e.g., Navarro, Frenk & White 1997). The mass of an NFW halo contained within a radius of r_{200} is therefore

$$M_{200} = M(r_{200}) = \frac{800\pi}{3}\rho_{\mathrm{c}}r_{200}{}^3 = \frac{800\pi}{3}\frac{\bar{\rho}(z)}{\Omega(z)}r_{200}{}^3 \tag{5.5}$$

where $\bar{\rho}(z)$ is the mean mass density of the universe at redshift z and $\Omega(z)$ is the density parameter at redshift z. It was shown that the characteristic overdensity is closely related to the halo's formation time: haloes which form earlier are more concentrated. Since more massive haloes assemble later, they are expected to be less concentrated, giving rise to an inverted concentration-mass relation. Consequently the total mass of a dark matter distribution can be calculated by integrating the local density over the radius r

$$M = \int_0^R 4\pi r^2 \rho(r) dr \tag{5.6}$$

As previously stated, the local value of the convergence may be expressed simply as the ratio of the local value of the surface mass density to the critical surface mass density. The radial dependence of the surface mass density of a spherically symmetric lens such as an NFW lens is obtained simply by integrating the 3-dimensional density profile along the line of sight

$$\sum(x) = 2 \int \rho(x,z) dz \tag{5.7}$$

with a dimensionless radial distance $x = \frac{r}{R_{\mathrm{s}}}$. The radial dependence of the convergence due to an NFW lens is then simply

$$\kappa_{\mathrm{NFW}}(\vec{\theta}) = \frac{\sum_{NFW}(\vec{\theta})}{\sum_{cr}} \tag{5.8}$$

Since the NFW density profile is spherically symmetric, the radial dependence of the shear can be written as

$$\gamma_{\mathrm{NFW}}(\vec{\theta}) = \frac{\sum_{NFW}^{-}(\vec{\theta}) - \sum_{NFW}(\vec{\theta})}{\sum_{cr}} \tag{5.9}$$

where $\sum_{NFW}^{-}(\vec{\theta})$ is the mean surface mass density interior to the dimensionless radius x.

5.2   Lensing map simulation

In order to properly simulate a model of weak lensing effects caused by dark matter, a weak lensing regime needs to be implemented on an otherwise random background of galaxies.
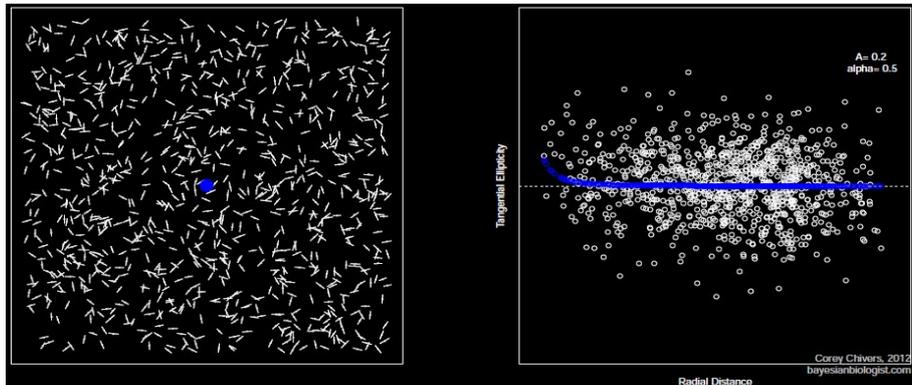
Figure 5.3: *Background galaxy lensing simulation. Credit: [35]*

By doing so, a simplistic 2D representation of the lensing influence in a otherwise 3D space can be displayed. The influences of both convergence and shear then become visible, depending on the amount of foreground mass and its distribution by local distortions.
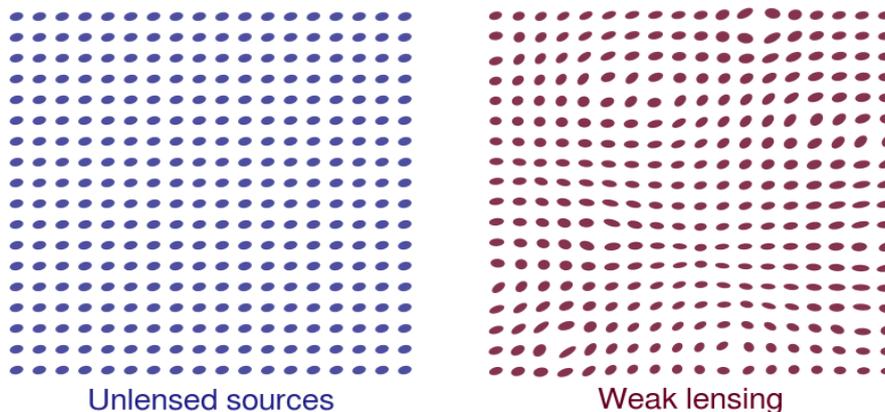
Figure 5.4: *Ellipticity distortion under weak lensing influence. Credit: Wikipedia*

N-body simulations are often used to create so-called shear maps. Derived from the gradient of the tangential alignment of galaxy filaments within a cluster, maps of the local apparent shear lead to the reconstruction of a convergence map. These are used to accurately calculate mass distributions of lensing objects, as can be seen below [36] [37] [38].
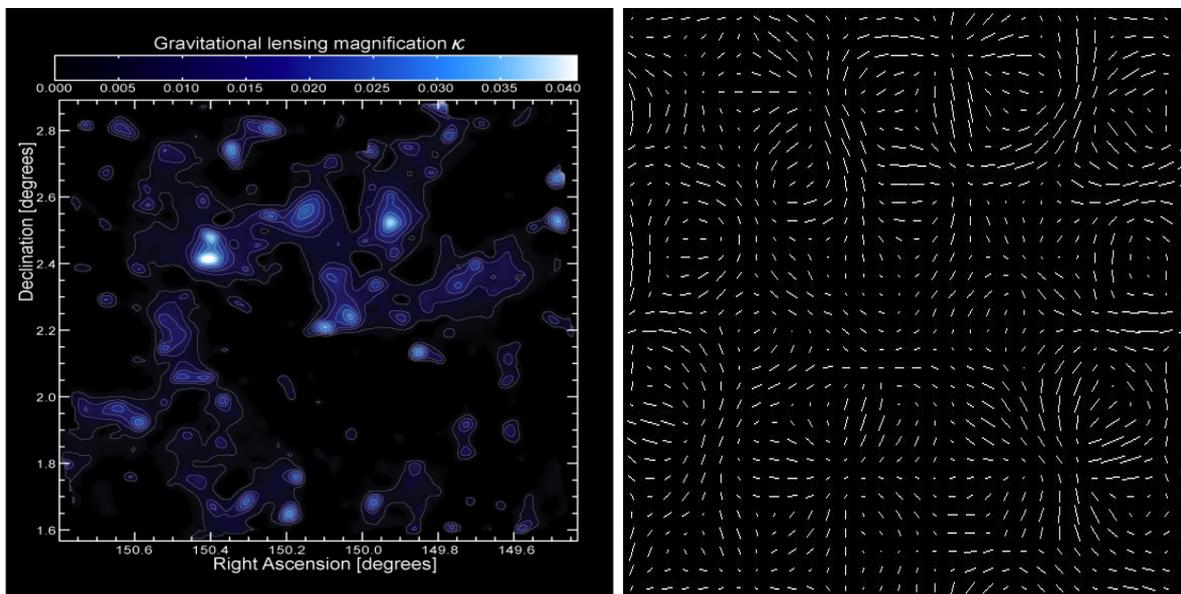
Figure 5.5: *Magnification map illustrating mass distribution. Credit: Cosmosat*

Figure 5.6: *Shear map, derived from ellipticity studies. Credit: PB works:gravity lensing*

In order to reduce the complexity of detecting lensing mass distributions based on optical analysis, the development of a training data set for a neural network application, will be based on recurring pattern formation, i.e. tangential shearing. As mentioned before tangential alignment of background mass consequently implies cross shearing of zero. Furthermore this approach supports lensing identification with convolutional layers.

5.2.1 Basic mathematical framework

To ensure consistency throughout the training data general mathematical conditions, describing the model, need to established. The gravitational potential from which the effective lensing potential is calculated for example, will be assumed to be generated by a point mass [17]. Thus the lensing potential can be solved for axial geometry

$$\Psi(\vec{r}) = \frac{D_{\rm ds}}{D_{\rm s}D_{\rm d}}\frac{2}{c^2}\int \Phi(\vec{r},z)dz = \frac{D_{\rm ds}}{D_{\rm s}D_{\rm d}}\frac{4GM}{c^2}ln(r) \tag{5.10}$$

with $\Phi = -\frac{GM}{R}$. Additionally the weak lensing regime usually features values of $\kappa \ll 1$ and $0 \leq \gamma \leq 1$, which is why the Jacobian matrix can henceforth be written with the unit matrix

$$A = \begin{pmatrix} 1-\gamma_1 & \gamma_2 \\ \gamma_2 & 1+\gamma_1 \end{pmatrix} \tag{5.11}$$

for $\kappa = 0$, with

$$\begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} = \frac{D_{\rm ds}D_{\rm d}}{D_{\rm s}}\frac{4}{c^2}\cdot\nabla\begin{pmatrix} \Phi_{11}-\Phi_{22} \\ 2\Phi_{12} \end{pmatrix} \tag{5.12}$$

and $\gamma_2 = 0$ for tangential alignment. Moreover the formula for γ_1 can be further simplified when the objects on the focal line of the observer are arranged to $D_{\rm ds} = D_{\rm d}$, with $D_{\rm s} = D_{\rm d}$:

$$\gamma_1 = \frac{2D_{\rm d}}{c^2}\cdot\nabla(\Phi_{11}-\Phi_{22}) = GM\frac{2Dd}{c^2}\cdot\left(\frac{1}{r_{11}{}^2}-\frac{1}{r_{22}{}^2}\right) \tag{5.13}$$

Alternatively, it can be written as:

$$\gamma_1 = r_{\rm s}\cdot D_{\rm d}\cdot\frac{1}{r_{\rm ij}{}^2} \tag{5.14}$$

The model assumes a variable Einasto profile of dark matter mass density

$$\rho(r) = \rho_0 \cdot exp(-Ar^\alpha) \tag{5.15}$$

with parameters A and α determining the strength of the tangential shearing caused by foreground dark matter [39]. That way the strength of the lens can be increased by varying the parameters of the Einasto profile as

$$M = 4\pi\int_0^R r^2\cdot\rho(\vec{r})dr = \rho_0\cdot\frac{4}{3}\pi r^{3-\alpha}\cdot exp(-Ar^\alpha) \tag{5.16}$$

As all background galaxies are simulated in an equal manor, the second term of formula 5.14 can be set to be a constant, describing either $D_{\rm d}$ or $D_{\rm ds}$ of all galaxies in respect to the optical axis. As the distance $D_{\rm d}$ is directly linked to the angle-dependent deflection regarding the locations $r_{\rm ii}$ and $r_{\rm jj}$, it needs to be determined while taking both prospects into account.

This way the requirement $0 \leq \gamma \leq 1$ can be adhered to. Now γ_1 solely depends on r_{ij}, which in a 2D depiction portrays the radial distance from the gravitational lense to the respective galaxy. Combined with formula 2.39, and the vector format in formula 5.12, the tangential rotation φ can be calculated as:

$$1 - \gamma_1 = cos(2\varphi) \Leftrightarrow \varphi = arccos(1 - \gamma_1) \tag{5.17}$$

With increasing shear, depending on mass of and distance from the lens, the rotation angle increases from $\varphi = 0°$ for $\gamma > 0$ up to $\varphi = 90°$ for $\gamma > 1$.

5.2.2 Cluster formation

Since a galaxy cluster is used as the source for the lens, its properties, as well as the properties of the dark matter halo contained within, need to be identified and determined. Even though galaxy clusters are the largest gravitationally-bound structures in the universe harboring star formations, black holes and AGNs[1], they are dominated by dark matter. Most of the normal light-emitting matter is in the form of hot plasma, the intracluster medium (ICM). Galaxy clusters cover a range of mass, with the lowest mass end being known to contain up to 100 galaxy groups [40].

	mass[%]	detection
Galaxies	1-2	Optical observations
ICM	10-13	Plasma at high temperature emits x-ray radiation
Dark Matter	85-90	Inferred through gravitational interactions

Table 5.1: *Relative proportion and detection of all mass in the universe.*

The ICM is hot due to the massive potential well of galaxy clusters. The gravitational potential energy of material falling into the cluster leads to shock heating of the gas to 10s of millions of °C or 2-15 keV, depending on the total mass of the cluster. Because of the size of clusters, this material is very tenuous, with only roughly between 10 and 10000 particles in each cubic metre of cluster. The density increases towards the centre of the cluster. The ICM emits X-rays strongly due to its high temperature via the Bremsstrahlung emission process. Supermassive black holes are often found at the centre of clusters, accreting matter which leads to highly energetic jets of material being accelerated into the ICM, preventing it from cooling rapidly. Even though clusters primarily consist of dark matter, it is evident, that its density is far from the critical *Schwarzschild density* ρ_c [41]:

$$\rho_c = \frac{3c^6}{32\pi G^3 M^2} = \frac{1.61 \cdot 10^{26}}{R_s{}^2} \frac{kg}{m^3} \qquad (R_s \ in \ meters) \tag{5.18}$$

Typically the mass of a cluster ranges from $10^{12} - 10^{16}$ solar masses ($M_\odot$). At the same time, diameters usually span from 1 to 5 mpc. Based on data collected from the milkyway and its nearest neighbour the *virgo cluster* 16.5 mpc away, measurements of the power spectrum were used to calculate the value of a quarter of the total energy density in the universe.

---

[1]Active Galactic Nucleus

Based on these findings, the dark matter lensing mass will be calculated in the following.

For spherical and radially symmetrical formations with $M = 10^{14} \cdot M_{\odot}$ and $d = 3 mpc$ the average dark matter density is found to be:

$$\rho_{\bar{\text{DM}}} = \frac{M}{V} = \frac{0.85 \cdot 10^{14} \cdot M_{\odot}}{\frac{4}{3}\pi \cdot (\frac{d}{2})^3} = \frac{1.691 \cdot 10^{44} kg}{4.155 \cdot 10^{69} m^3} = 4.069 \cdot 10^{-25} \frac{kg}{m^3} \tag{5.19}$$

The literature value is found to be slightly higher at $\rho_{\text{DM}} = 8.913 \cdot 10^{-24} \frac{kg}{m^3}$ [42].

This approximation also checks out when integrating the core density ρ_0 in regard of the Einasto profile with the variables:

α	A	r [Mpc]	$\rho_0[\frac{kg}{m^3}]$	M [kg]	r_s [pc]
0.06	0.2	1.5	10^{-21}	$1.858 \cdot 10^{44}$	8.942

Table 5.2: *Example parameters for dark matter distribution in galaxy cluster.*

Note that the core denisty ρ_0 is larger than $\rho_{\bar{\text{DM}}}$, as the cluster density, e.g. the dark matter density increases inwards.
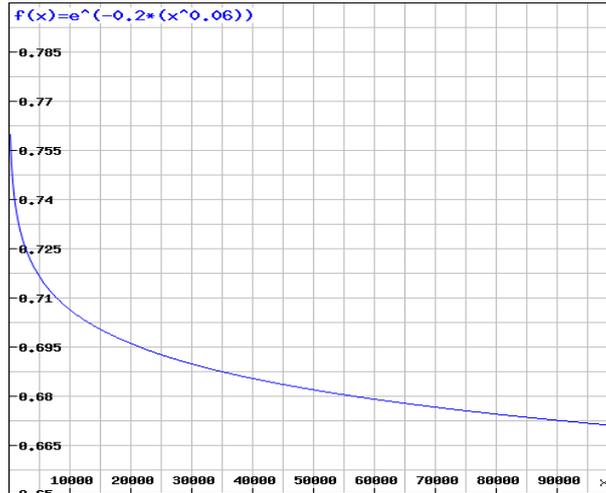
Figure 5.7: *Dark matter density function. Credit: Rechneronline.de*

For a set lensing mass M_{DM}, the tangential shearing can be calculated considering only the lensed galaxy span r_{ij} in regard to the breadth D_{ds}. As mentioned before, the relation of these two variables needs to meet the condition of $0 \leq \gamma \leq 1$. As later simulations will set a minimal distance at $0 \leq r_{ij}$, the range will be fit to a [-8, 8] xy-grid in units of mpc for a fixed galaxy plane located 10 Gpc away. That way formula 5.13 is simplified to:

$$\gamma = \frac{2GD_{ds}}{c^2} \cdot \frac{M_{\text{DM}}}{r_{ij}^2} = 4.813 \cdot 10^{-46} \frac{mpc^2}{kg} \cdot \frac{M_{\text{DM}}}{r_{ij}} \tag{5.20}$$

The following simulations will apply a mathematical torque on the elongated galaxies, rotating them to generate circular tangential alignment around the lensing mass. The then visible effect will decrease with increasing distance. Thus the pattern in regard to the placement and number of foreground masses, can be used to train the optical neural network.

5.3 Data creation in Python

In similar fashion to the left image in figure 5.3 the training data consists of a square frame filled with short dashes each with a centered rotational axis. These dashes, representing the background galaxies, are scattered at random while being distributed uniformly. When scattered, the additional degree of freedom regarding the rotational orientation is also assigned arbitrarily. After placing a mass on the 2D plane, the otherwise expected tangential shearing of the background galaxies will be converted to a rotational angle. That way the intrinsic galaxy ellipticity can be displayed without having to induct γ into the dimension of the semi-axis. The focus will be set solely on the spin of the galaxies. This tangential alignment will then be made visible, by applying a angle of rotation to all galaxies in relation to their placement on the 2D plane. In addition, the galaxies will be stretched tangentially around the foreground mass, illustrating the effect of arc emergence. The data created will then be saved in a frame which can later be used to be fed into the neural network.

5.3.1 Lensing plots

The plots which later will serve as training and validation data are created in an open source cross-platform integrated development environment (IDE). In this case, the Anaconda IDE Spyder. First, the necessary libraries need to be imported.

```python
# @author: William Roster

import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import pickle
from PIL import Image, ImageDraw
```

The background galaxies are generated with random coordinates and orientations as well as a random intrinsic ellipticity. Later is measured in regard to a tangent on the equipotential line engulfing the lensing mass. Simultaneously the cross product between the galaxy vector itself and the vector spanning from lensing mass to the galaxy serves as placement detection and allows for a establishment of the correct adjustment direction as each quadrant needs to be handled differently.

```python
def generate_arrow_array(size, massp = [0,0]):
    # initializing variables
    angle = [], dx = [], dy = [], x = [], y = [], c = [], c_start = []
    midx = [], midy = [], vecMGx = [], vecMGy = [], relAngle = [], dist = []

    # calculate intrinsic ellipticity
    def angleCalc(vec1, vec2):
        skal = vec1[0] * vec2[0] + vec1[1] * vec2[1]
```

40

```python
9          absV = np.sqrt(vec1[0]*vec1[0] + vec1[1]*vec1[1]) * np.sqrt(vec2[0]*vec2[0] +
              vec2[1]*vec2[1])
10         tempAngle = np.arccos(skal/absV) #cutting angle
11         if(vec1[0] * vec2[1] - vec1[1] * vec2[0] > 0):
12             if(tempAngle > np.pi/2):
13                 tempAngle = np.pi - tempAngle
14                 orient = 1 # cross product serves as indicator for later adjustment
                      turning direction
15             else:
16                 orient = -1
17         else:
18             if(tempAngle > np.pi/2):
19                 tempAngle = np.pi - tempAngle
20                 orient = -1
21             else:
22                 orient = 1
23
24         return (np.pi/2 - tempAngle),orient #intrinsic angle in relation to
              equipotential
25
26     # generate random orientations and rotations
27     for i in range(0, size):
28         angle.append(np.random.rand()*np.pi)
29         dx.append(length*np.cos(angle[i])*8)
30         dy.append(length*np.sin(angle[i])*8)
31
32         # generate random coordinates
33         x.append((np.random.rand()*2-1)*8)
34         y.append((np.random.rand()*2-1)*8)
35
36         # calculate rotational axis
37         midx.append(x[i] + dx[i]/2)
38         midy.append(y[i] + dy[i]/2)
39
40         # vector stretching from lensing mass to galaxy
41         vecMGx.append(midx[i]-massp[0])
42         vecMGy.append(midy[i]-massp[1])
43         dist.append(np.sqrt(vecMGx[i]**2 + vecMGy[i]**2))
44
45         # convert intrisic rotation to value[0,1]
46         for i in range(0, size):
47         c_start.append(angleCalc((vecMGx[i],vecMGy[i]),(dx[i],dy[i]))[0]/1.57)
```

As described in subsection 5.2.1 the rotation angle is derived from γ, which
in turn depends on the mass as well as the relative distance. The borderline case is set at
$\gamma = 1$, as more than perfect alignment is not achievable. Thus the intrinsic rotation sets a
limiter on the applied rotation. All calculations are made in the scale of mpc.

```
1  def gamma(d,m,g,c0,mpc):
2      rs = []     # Schwarzschild radius
3      ds = []     # distance in image plane
4      rsmpc = [] # Schwarzschild in mpc
5      rs = ((2*g*m)/c0**2)
6      rsmpc = rs/mpc
7      ds = 10000 # 10 gpc
8      gammaE = (((rsmpc*ds)/(d)**2)) # see formula 5.14
9      if (gammaE > 1):
10         gammaE = 1 # borderline case
11
12     return gammaE
13
14  def phi(d,m):
15
16     return np.arccos(1-gamma(d,m,g,c0,mpc)) # applied rotation angle, see formula 5.17
```

In the following, φ is applied to all background galaxies. Thus all coordinates, orientations and angles need to be recalculated and adjusted. Another feature of lensing is obviously illustrated by the tangential smear of the galaxies closest to the lensing mass. Consequently the galaxies are stretched in regard to the value of φ. This can be seen clearly when comparing figures 5.8 and 5.9.

```
1  def update_arrow_array(midx,midy,dist,angle,relAngle,mass, size):
2      newAngle = [], newx = [], newy = [], newdx = [], newdy = []
3
4      for i in range(0, size):
5          tmpAngle = phi(dist[i],mass)
6          stretch = 1 + (tmpAngle * 2) # tangential smear
7          if(tmpAngle > relAngle[i][0]): # rotation limiter
8              tmpAngle = relAngle[i][0]
9          if(relAngle[i][1] == -1): # rotation adjustment direction
10             tmpAngle = -tmpAngle
11         newAngle.append(angle[i] - tmpAngle) # resulting ellipticity
12
13         # new mass to galaxy vector
14         newdx.append(length*stretch*np.cos(newAngle[i])*8)
15         newdy.append(length*stretch*np.sin(newAngle[i])*8)
16
17         #new galaxy coordinates
18         newx.append(midx[i]-(newdx[i]/2))
19         newy.append(midy[i]-(newdy[i]/2))
20
21          # convert applied rotation to value[0,1]
22          for i in range(0, size):
23         c.append(phi(dist[i],mass)/1.57)
```

A panda data frame is used to properly observe and monitor the adjustments of all the galaxies and their characteristics. The data frame can either monitor every individual galaxy in regard to their coordinates and rotation before and after lensing or document the lensing mass for every example generated. For later, the range is set to the variable "example".

```python
def gen_data_frame_for_plot(midx, midy, dist, angle, relAngle, mass, massp):
    data = {'midx' : midx,
            'midy' : midy,
            'dist' : dist,
            'angle' : angle,
            'relAngle' : relAngle,
            'mass' : [mass for i in range(0, len(midx))],
            'masspx' : [massp[0] for i in range(0, len(midx))],
            'masspy' : [massp[1] for i in range(0, len(midx))]}

    df = pd.DataFrame(data, columns=['midx', 'midy', 'dist', 'angle', 'relAngle',
        'mass', 'masspx', 'masspy'])

    return df
```

Finally the programme can be run in order to generate training data by applying a random mass in the range of $10^{44\text{-}45}kg$ at a random location. The plots are then saved in numbers up to 10s of thousands to provide an extensive training and validation data set. In this case 750 background galaxies are chosen, to avoid creating a fluid-like dynamic, which would be too easy to spot, while ensuring enough structure for pattern formation to commence.

```python
for j in range(0,example):     #  example amount of random data sets
    # generate random mass and mass coordinates
    random_m = np.random.rand()*10
    if(random_m < 5):
        mass = ((random_m+5)*10**44)
    else:
        mass = ((10-random_m)*10**45)
    massp=((np.random.rand()*2-1)*8),((np.random.rand()*2-1)*8)
```

One example of such a lensing simulation can be seen in figures 5.8 and 5.9. The color scheme in both figures shows the absolute ellipticity. As at first the intrinsic ellipticity is the only influential parameter, the full scheme of the colorbar is spread uniformly across the 2D plain. After the lensing effect takes places, the greatest ellipticities are found closest to the lensing mass. The data used later to train and validate the neural network will be of simplified nature. As the intensity of tangential shearing is not made visible by a color scale in space, the neural network will be working with black and white binary images. That way the network will only be training in a 1 dimensional color scheme, thus reducing the necessary amount of data.
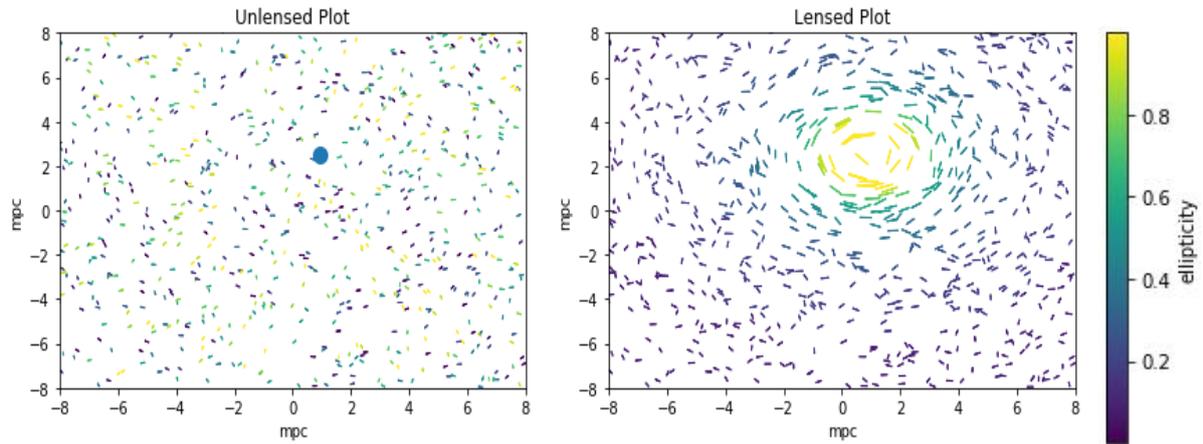
Figure 5.8: *Uniformly distributed galaxies without the influence of a foreground lensing mass illustrated by a marker. (unlensed)*

Figure 5.9: *Direct comparison to fig. 5.8 under the influece of m = 3.377046392177419e+45 kg. The relative ellipticity is illustrated with the respective color scheme. (lensed)*

5.3.2 Input conversion

Furthermore the plots will be saved at a resolution of 200x200 pixels. With features taking up approximately $\frac{1}{16}$ of the 2D plane, this results in exactly 2500 pixels per event. In comparison, the well known MNIST database of handwritten digits with black and white numbers scaled at 28x28 pixels result in 784 pixels per image. This way the network will be able to detect even the smallest of events in high resolution wide angle recordings.

```python
img2 = Image.new("L", (pix,pix))
draw = ImageDraw.Draw(img2)
pixel_coords_x = pixel_mpc*df_final[j].x.values + (pix/2) #coordinates to pixel
pixel_coords_y = -pixel_mpc*df_final[j].y.values + (pix/2)

for i in range(0, (len(pixel_coords_x))): #set limiter to 200x200pixel
    if (df_final[j].xdx[i] > 8):
        df_final[j].xdx[i] == 8
    elif (df_final[j].xdx[i] < -8):
        df_final[j].xdx[i] == -8
    elif (df_final[j].ydy[i] > 8):
        df_final[j].ydy[i] == 8
    elif (df_final[j].ydy[i] < -8):
        df_final[j].ydy[i] == -8

pixel_coords_xdx = pixel_mpc*df_final[j].xdx.values +(pix/2)
pixel_coords_ydy = -pixel_mpc*df_final[j].ydy.values + (pix/2)
for i in range(0, (len(pixel_coords_x))):
    draw.line((( pixel_coords_x[i] , pixel_coords_y[i] ),( pixel_coords_xdx[i] ,
        pixel_coords_ydy[i] )), fill='white', width = 1) #draw line for lensed galaxies
```

For the network to be able to digest 40000 input values, the images need to be broken into single pixel values and then converted into a 2D numpy string. This is done for both the labeled and the unlabeled input data. Here 200x200 images are generated and then drawn upon with the same data available as displayed in images 5.8 and 5.9. Afterwards they are appended to an array which is later normalized and 0-centered before being converted to a .npy file and saved on the disc. The generated data is now seperated in 2 arrays. The first array holds a list of "example-amount" 2D numpy arrays broken down into pixel values of either -0.5 or 0.5. This data will serve as image input. The second array is a list of the same amount featuring 3D vectors holding the normalized pixel values of the mass coordinates and the normalized value of its weight. Their values range from 0 to 1 as this has to do with later calculations within the network, including the activation function and will be used as the reference data.

```
1  # save image data to numpy array - list of np arrays
2
3  img2.save(path3 + str(j) + path4)
4  arr = np.array(img2)
5  wholeArray.append(arr)
6  wholeArray=np.asarray(wholeArray)
7  wholeArray_labelled=np.asarray(wholeArray_labelled)
8  divider = [255]
9  wholeArray=wholeArray/np.max(divider) - 0.5
10
11 #save arrays as list
12 np.save(path5 +"Unlabeled_"+"("+str(len(wholeArray))+",200,200)" , wholeArray)
13 np.save(path5 +"triple_Labeled_"+"("+str(len(wholeArray_labelled))+",1,1,1)" ,
        wholeArray_labeled)
```
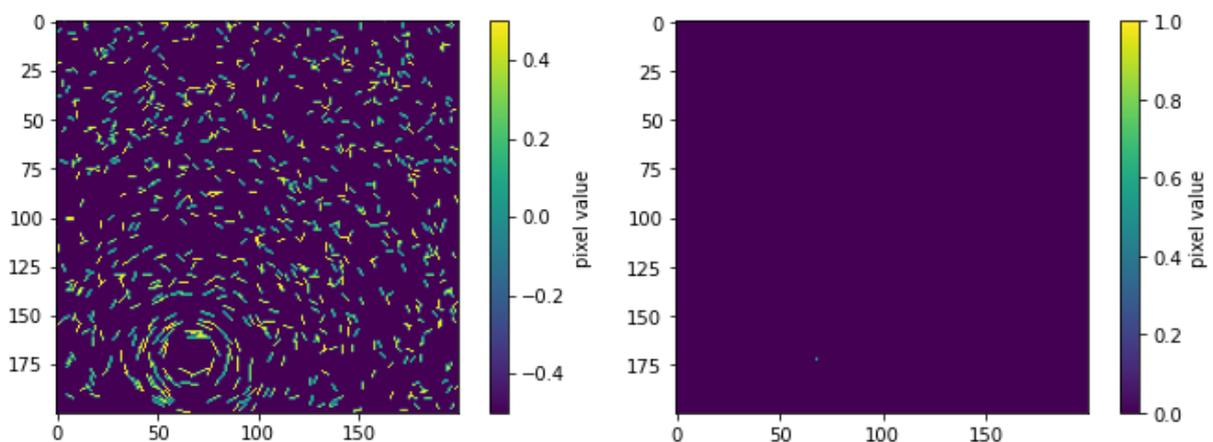

Figure 5.10: *Uniformly distributed galaxies under the influence of m= 2.8297352833734411e+45 kg, imported as an array figure. Resized to image format 200x200 pixel.*

Figure 5.11: *Input reference data to fig. 5.10 holding information about the mass coordinates at [y= 173, x=66].*

These images encoded in the 2D arrays can then be imported and re-displayed as figures in the neural network. An example is given above. In this case the reference input is also portrayed as an image, however an image as reference for an image would not work, as the network would have to be able to find the right pixel among 40000. That is why the coordinates can either be converted into values ranging from 0 to 200 or normalized as recently explained, drastically reducing the computing complexity.

Chapter 6

Neural network

Even though the primary task of a Convolutional network is ususally known to be performing image classification, the field of computer vision has been revolutionized since the introduction of CNN's for two simple reasons. Firstly images used for computer vision problems nowadays are often 224x224 or larger. That comes out to 224 x 224 = 50,176 input features. A typical hidden layer in such a network might have 1024 nodes, so 50,176 x 1024 = 51.38 million weights and biases would have to be trained for the first layer alone. The seemingly simply task can then quickly become extremely extensive and impossible to implement. Luckily, pixels are most useful in the context of their neighbours. As images are made up of objects featuring small, localized features the necessity for every node to look at every pixel in the first hidden layer becomes redundant. Secondly, the locations of features within an image vary. However, the network should be able to detect features, regardless of their position. This chapter will therefore focus on the step by step implementation of a convolutional neural network.

6.1   Implementing a CNN in Keras

As mentioned before Keras is a high level API which allows users to develop and design neural networks on a tensorflow backend. As tenserflow is a low level API implemented in Python the following source code is also written, compiled and executed in the Anaconda IDE Spyder. Similar to the creation of the training data the necessary libraries need to be imported first. In doing so, the respective tools used to build the model are also imported.

```
1  # @author: William Roster
2
3  import tenforflow as tf          #backend
4  import numpy as np
5  from numpy import load
6  import matplotlib.pyplot as plt
7  import keras.backend as K
```

It is a known fact, that Keras performs best when the input is 0-centered. Additionally Keras expects a 4 dimensional input featuring the amount of data, its 2D pixel values and the number of features to expect, in this case 1 (binary). After loading the input data for training and validation the images need to be reshaped to fit Keras demands.

```
1   #import normalized data from numpy arrays
2
3   train_images = load("/Users/william/Desktop/PYTHON/numpy_arrays/
4             np_array_Unlabeled_(50000,200,200).npy")
5   train_all_labels = load("/Users/william/Desktop/PYTHON/numpy_arrays/
6             np_array_triple_Labeled_(50000,1,1,1).npy")
7   test_images = load("/Users/william/Desktop/PYTHON/numpy_arrays/
8             np_array_Unlabeled_(10000,200,200).npy")
9   test_all_labels = load("/Users/william/Desktop/PYTHON/numpy_arrays/
10            np_array_triple_Labeled_(10000,1,1,1).npy")
11
12  #resize images
13  train_images = np.expand_dims(train_images, axis=3)
14  test_images = np.expand_dims(test_images, axis=3)
15
16  print(train_images.shape)          # (example, 200, 200, 1)
17  print(train_all_labels.shape)      # (example, 3)
18  print(test_images.shape)           # (example2, 200, 200, 1)
19  print(test_all_labels.shape)       # (example2, 3)
```

Next the filters and pooling layers are defined. These are called upon, when the network layers themselves are filled with parameters. Even though the network designed, works in a sequential order it is set up using Keras functional API model rather than the sequential model for reasons of flexibility. The functional API offers many alternatives, such as shared layers, multiple inputs and parallel motions. This allows for the network to be altered in any desired way. More complex network architectures usually perform better with more extensive problems featuring further unknown variables.

6.1.1 Functional API Model

After setting up the layers with defined inputs and outputs the model can be constructed and viewed using summary. Now the model can be compiled. Here the optimizer and its learning rate as well as the loss function and the visualization are defined. Then the model is fitted with training and validation data.

```
1   def euclidean_distance_loss(y_true, y_pred): # loss function
2       return K.sqrt(K.sum(K.square(y_pred - y_true), axis=-1))
3
4   model.compile(tf.keras.optimizers.Adam(lr = 0.0008), loss=euclidean_distance_loss,
5           metrics=[tf.keras.metrics.RootMeanSquaredError()],)
6
7   history = model.fit(x=train_images, y=train_all_labels, epochs=40,
8     validation_data=(test_images, test_all_labels), shuffle=True, batch_size= 64)
9
10  model.save_weights('cnn_final.h5')
11  model.save("triple.h5")
```
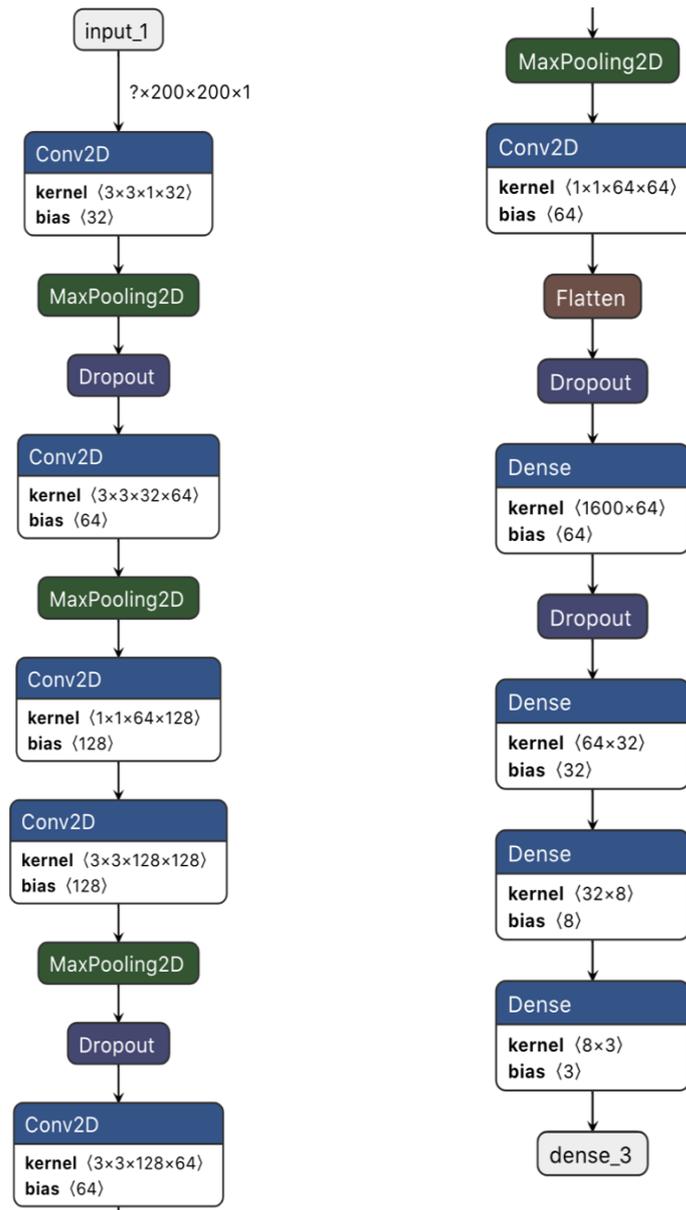
Figure 6.1: *Visualization of the program flow chart in sequential design. Each layer contains further information about the respective parameters being used. Credit: Netron*

After the batch size is set the network will start training and backpropagating for the number of epochs set before. This way the network can be called upon when wanting to apply it to a new set of data without having to run the training sequence once more. When calling summary, python displays an overview of all the network layers and the individual parameters featured within each layer, as well as the total number of trainable nodes and biases. This can be illustrated using *Netron*, see above. The returned parameter overview is displayed in the list below:

- Total parameters: 357,507

- Trainable parameters: 357,507

- Non-trainable parameters: 0

6.2 Network validation

After supplying the model with all necessary input the training routine is run for the chosen number of epochs. The number of epochs is crucial, as the network should be trained enough to have a solid level of accuracy while not being overfitted at the same time. This is usually prohibited from happening by adding so called *Dropout* layer in the networks architecture. These are handed a parameter ranging from 0 to 1 representing the percentage of nodes which are randomly ignored when running a training sequence. This effect also allows the network to learn more selectively with deeper features hidden within the convolutional layers. Depending on the optimizer options such as the "slow gradient descent" (SGD) feature a manually settable learning rate (LR). Again, this involves a little bit of trial and error as there is no universal applicable "best setting". If the learning rate is too low the network might get stuck in a local minima, if it is too high it might keep overshooting the global minima instead of gradually moving closer and finally converging.

While training, the network continuously computes the MSE or, put simply, the loss as well as the inverse accuracy for both, the training and validation set after every epoch. The time needed to train the network on 1 epoch depends on the number of trainable parameters, as backpropagation is performed after every training circle. This is also why more often than not, less is more. A network with millions of connections might have trouble finding decisive advantages with changing specific weights for certain nodes. That being said the network might need plenty connections which all allow to be altered slightly to perform better results. Usually a larger set of training data with more complex patterns requires more network parameters. After training is completed the training results can be displayed. The validation results obviously serve as benchmark for further external data.

6.2.1 MSE results

The network performance after training completion is displayed in the table below. Obviously the end result is completely dependent on all parameters and will most definitely perform distinctively different under new settings. Thus far the following parameter setup allowed the network to output the best training results.

input$_{train}$	input$_{val}$	loss$_{val}$	MSE$_{val}$
50000	10000	0.0379	0.0301
epochs: 40			

Table 6.1: *Network parameter output and summary*

As can be seen in the figures 6.2 and 6.3 the deviation error is marginal as of the first epoch. Afterwards the loss as well as the MSE only gradually decline with increasing epochs. Higher percentages, i.e. steeper climbs or slower, less steep approaches could be achieved by alternating the framework of the network by changing the number of parameters involved or whole layers themselves.
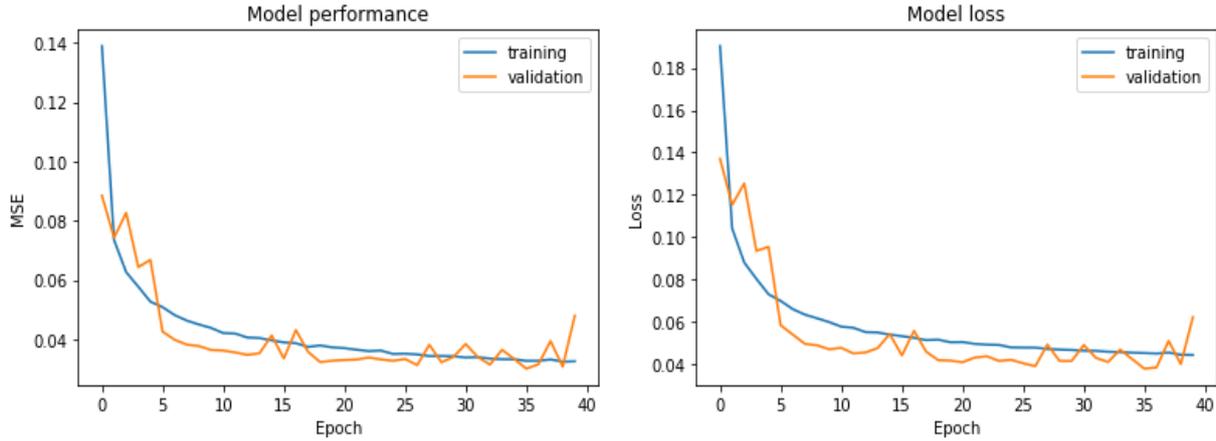
Figure 6.2: *Plot illustrating both, the training and validation mean squared error (MSE) for a training sequence 40 epochs.*

Figure 6.3: *Plot illustrating both, the training and validation model loss, calculated in euclidean vector lengths, for a training sequence 40 epochs.*

In this case the network achieves a validation accuracy of roughly 96.5%. Even though this is a rather good result it needs to be taken into account, that the data used, was normalized. This means that the values ranging from 0 to 1 obviously represent percentages. A second percentage of a first percentage is therefore strongly dependent on the first value.
For example 3% error of 0.4 of a coordinate equals: $\pm$ 0.03 $\cdot$80 $=\pm$ 2.4.
On the other hand 3% error of 0.8 of a coordinate equals: $\pm$ 0.03 $\cdot$160 $=\pm$ 4.8. This implies that the networks deviation needs to be interpreted under the observation whether the result is of high >1 or low >0 nature. In addition the graphs illustrating the validation error and loss are very unsettled. This might be due the fact, that the learning rate is too low and the network performance is stuck in a local minima. At the same time a lower learning rate could smooth the validation graph, producing results with more continuity and stability throughout the training epochs. However the training graphs are very steady and constantly decreasing in comparison. This can be a case of overfitting which could be solved by introducing regularisers to the layers. On the other hand is may be plausible that the network needs more epochs to gradually converge and balance out the irregularities.

6.3 Cross-Check performed on Universe footage

As the network focuses on pattern recognition only, black and white binary images of the night sky, capturing weak lensing effects, should serve as an ideal test for final validation. As the network has only "known" binary data, any test data needs to be transformed in to according format first. Luckily any rgb image can first be turned into a grayscale image before being further reduced in features by applying a threshold for black and white class segregation. This does however also harbour error potential as galaxies and in particular galaxy clusters are not evenly bright. They usually have a brightness distribution which depends on the amount of stars at a specific location within the cluster and decreases outwards. When turning the rgb image to grayscale the according colours are interpreted differently.

This means that the color of the emitted light, which ultimately might be evenly bright, is valued differently. After the threshold is applied to turn the images into binary, the before feature heavy grayscale image is converted into a hard edge blur of black and white. This has great influence on the networks performance as extremely elongated galaxies, which serve as best possible lensing indicator are dropped, as they appear less bright than an ordinary galaxy. That way the images losses important information about the most telling section of the image. One potential solution to this problem would be to use grayscale simulations. However these require a closer-to-reality image, which would have to be copied from deep space recordings. As the network needs extensive amounts of training data the generation of such simulation data would be very tedious.

6.3.1 Ratification with Hubbel

Deep filed images taken by the hubble space telescope have been giving astronomers a much better understanding of the universe for the past 30 years. Among the pictures taken, galaxy clusters are also often found. Even though the images are usually taken so that the main attraction, the centre of the cluster is in the middle of the picture, these images can still be used as proper validation for the network. Besides the estimate for the DM coordinates, the network will also calculate a value for the mass. Obviously this value is restricted to the mathematical framework introduced in chapter 5.2.1 which accurately states the region of mass and space taken into account. While the coordinates are independent and can be approved by looking at the image there is no legit measurement of the dark matter distribution among either cluster. In the following two validation examples will be given.

- **Abell 1689**

  One of the best known and studied galaxy clusters in the greater research area of dark matter accumulations is the cluster named Abell. It has many different sections, each accounted for with a number tag. In this case the network will be looking at an image taken from abell 1689 which shows intense weak lensing effects as can be seen in the sectional cutout in fig. 6.5.

Figure 6.4: *Abell 1689 and cutout section under the influence of weak lensing. Credit: NASA*

First of all the image needs to be transformed as explained above. The network then predicts the coordinates [y: 97, x: 110]. This pixel and the surrounding area is highlighted in red. Keeping the validation error of 0.03 in mind, this calculation results in y: 97 ± 2.91, x: 110 ± 3.30. With a root MSE of approximately 7.3 pixels this equals a deviation of 4.09%, making this calculation 95.91% accurate.

This result is based on my personal assumption, that the mass is placed at [y: 90, x: 112]. The calculated mass is illustrated by the yellow dot, circled red, wheres as my human suggestion is marked red, circled white.

Figure 6.5: *Binary conversion of fig. 6.4(left) and respective mass coordinates(right) highlighted for both, network and human estimate. Credit: NASA*

- **MACS J0416.1-2403**

  This is one of six galaxy clusters being studied by the Hubble Frontier Fields program, which together have produced the deepest images of gravitational lensing ever made. It is thought that MACS J0416.1-2403 contains a significant amount of dark matter. Compared to other commonly observed galaxy clusters, it is more efficient at producing multiple lensed images of background galaxies [43].
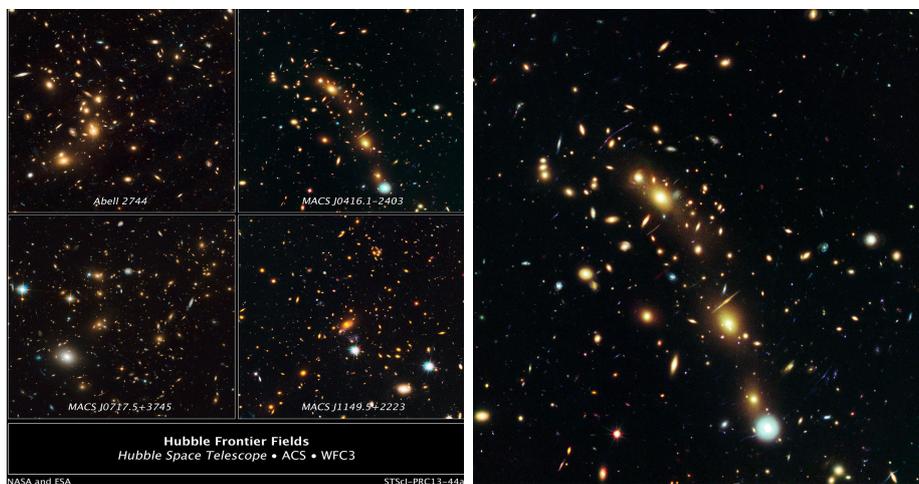
Figure 6.6: *MACS and cutout section J416 under the influence of weak lensing. Credit: NASA*

This has primarily come into focus after it was announced as gravitationally lensing the most distant galaxies at redshift z=12. In this second example the network outputs the coordinates [y: 69, x= 59]. The error remains the same, thus the error margin is y: 69 $\pm$ 2.07, x: 59 $\pm$ 1.77. For an estimated mass at [y: 67, x: 63], the absolute deviation is 4,2 pixels or 3.54%. This, in respect of the pixel values, converts to a accuracy percentage of 96.46%.
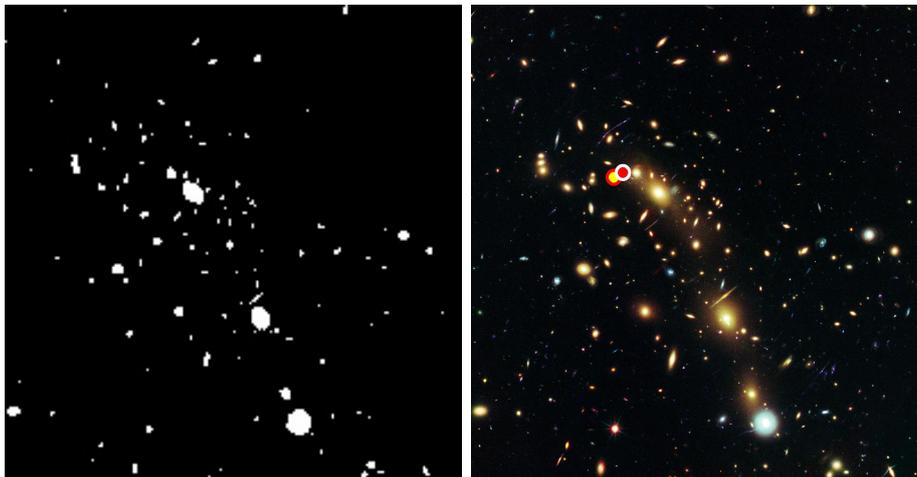
Figure 6.7: *Binary conversion of fig. 6.6(left) and respective mass coordinates(right) highlighted for both, network and human estimate. Credit: NASA*

6.3.2 Evaluation

For both examples, the network predictions were almost perfect. These estimates obviously depend on where the mass distribution is actually placed in reality. In this way and form one can only assume by investigating and interpreting the lensing patterns among the images. In addition dark matter halos or accumulations are usually spread throughout the cluster, turning a pin-point location into an educated guess. However, the network did perform extremely well under the conditions given. If further adjustments, regarding the input data pre-processing, were to be made, neural networks certainly provide the capability of reliably producing high-quality outputs for given mathematical frameworks, exceeding beyond 96% accuracy.

Chapter 7

Conclusion

With artificial intelligence being implemented ever deeper in day to day life, this thesis attempts to apply modern data science algorithms to astronomical forefront research in the field of dark matter. With a continuously growing understanding of the universe and its composition, dark matter has recently become one of the biggest questions to be answered when it comes to solving complex theories such as the accelerating expansion of the universe. Unfortunately the examination of dark matter is vastly restricted, begging for alternative study methods.

The progress made in the field of data analysis in recent times, in particular when it comes to object recognition, allows for scientific research to process and handle experimental data in new ways. Ultimately acquiring knowledge from correlations undiscovered before. A similar approach was taken in the frame of this thesis as artificial intelligence is trained to discover and characterize dark matter accumulations in space. These are currently known to predominantly appear in galaxy clusters. However, the experimental data on dark matter is minimal and neural networks require thousands of training data examples in order to properly apply weights and biases to achieve acceptable results. This is why, based on general dark matter knowledge, a frame of appearance for a mathematical description was defined in order to generate simulated lensing data.

For reasons of simplicity and to force the network to look for hidden features amongst the training data the images were uploaded as binary pixel arrays. With neural networks being multi functional many forms of learning architectures can be implemented. These include MLP's, Regression and Categorization. The network presented in this thesis is of strictly sequential manor. By doing so, both the coordinates as well as the mass are estimated for a given image. Consequently the loss function considers all 3 outputs when altering the system variables. However this can lead to inaccuracy as the network can use convolutional layers to filter the image in order to find the lensing pattern. The mass is therefore also depicted from the whole image, even though its value prediction is strongly linked to the magnitude of the tangential shearing in immediate proximity around the mass coordinates. To overcome this issue it is highly recommendable to split the sequential network model, thus allowing it to use two inputs instead of one. This way the network would first use the simulated images to establish the mass coordinates. These coordinates would then be forwarded into a secondary input layer, which would also be fed with the, in the first input layer featured, simulated images.

From this secondary input layer the network can then perform regression, by finding a correlation between the location and the pattern before starting to make predictions. This way the network would learn to concentrate on the location and therefore the desired section of the image, before any further calculations were to take place. Besides the alternation of the number of inputs, making use of the functional API also enlarges the possibilities of multi-layer connections such a used in ResNet. Thereby multiple connections can be strung without overwhelming the network with millions of system parameters. In general the focus should be set on making the network deeper, by adding many small layers, rather than wider by cramming more information into bigger filters.

Future analysis regarding this topic should also consider implementing the possibility of multiple lensing effects at once. As mentioned before, dark matter is spread across galaxies and often tends to be located on the outer edges, rather than centered in the middle. X-ray surveys may deliver more applicable data to properly simulate dark matter density distributions in space in order to properly identify accumulations in clusters when observing weak lensing events. In conclusion computer vision tools, such as pattern recognition, prove to be implementable for any given object recognition task under the prerequisite of well prepared and purified data, as networks are only as smart as one programmes them to be.

Appendices

Appendix A

Hard drive: Python documentation

The documentation of the code written and used in this thesis for:

- The generation of training and validation data sets on the creation of weak lensing patterns

- The design of a neural network performing weak lensing pattern recognition

can be found on the hard drive CD which is attached to the inside rear cover of the thesis.

Acknowledgements

The outbreak of the corona pandemic in the spring of 2020 has made the elaboration of this thesis especially challenging. This is why I want to take the time to thank all the people who have been particularly helpful and supportive for the past few months.

First, I would like to thank PD Dr. Jens Sören Lange for all the advice, time, and sincere conscientiousness. He has been an excellent advisor. When ever I would reach a dead end, he would be able to point me in the right direction, while never condemning my scientific approach to problems. Klemens and Eli, as well as Janik and Marvin for helping me out with any sort of programming issues, thus overcoming one of the more tricky parts of writing this thesis. Stephanie for proof-reading my thesis as well as the rest of the research group, for kindly providing me with useful knowledge when seen fit or requested upon my behalf. Not to forget, I owe a great debt of gratitude to PD Dr. Sophia Heinz for being my second reader.

Although it is difficult to list them, I would like to thank all my friends in Giessen, especially: Eli, Marian, Janik and Zoe for always sticking up for me since the beginning of the degree. They have turned day to day university life in to fun and I have learned something from all of them. In addition I would like to share my gratitude for all of you having to endure my talks about dark matter and neural networks for the past few months.

Lastly I would like to say big thank you to my family. Everyone has been extremely supportive over the years, always making sure I was moving in the right direction. In contrast, the large amount of laughter and sarcasm which runs through the family has definitely helped me to never see helpless situations too serious.

Finally I would like to thank both, the JLU and THM and all those who were involved, during the countless hours spent in achieving the goal of a Bachelors degree in physics. And so the journey continues.

Giessen, 2020

William Thomas Roster

List of Figures

List of Tables

Abbreviations

AGN	Active galactic nucleus
AI	Artificial intelligence
ANN	Artificial neural networks
API	Application programming interface
BAO	Baryon acoustic oscillation
BNN	Biological neural network
CDM	Cold dark matter
CMB	Cosmic microwave background
CNN	Convolutional neural network
CPU	Central processing unit
DEE	Distributed execution engine
DM	Dark matter
DNA	Deoxyribonucleic acid
EM	Electromagnetism
ENIAC	Electrical numerical integrator and calculator
ESA	European space agency
FBNN	Feedback neural networks
FFNN	Feedforward neural networks
GIMP	Gravitationally interacting massive particle
GPU	Graphic processing unit
GR	General relativity
ICM	Intracluster medium
IDE	Integrated development environment
LAT	Large area telescope
LR	Learning rate
LSP	Lightest supersymmetric particle
LUX	Large underground xenon experiment
MLP	Multilayer perceptron
MNIST	Modified national institute of standards and technology database
MOND	Modified newtonian mechanics
MSE	Mean squared error
NASA	National aeronautics and space administration
NFW	Navarro-Frenk-White
ONEIROS	Open-ended neuro-electronic intelligent robot operation system
ResNet	Residual neural network

SGD	Slow gradient descent
SM	Standard model
TLU	Threshold logic units
TPU	Tensor processing unit
VERITAS	Very energetic radiation imaging telescope array system
WIMP	Weakly interacting massive particle
WMAP	Wilkinson microwave anisotropy probe

Bibliography

[1] B. Cox, A. Cohen. *Human Universe & Forces of Nature*, William Collins 2017, BBC 2014.

[2] B. Cox, A. Cohen. *Wonders of the Solar System and the Universe*, William Collins 2015, BBC 2014.

[3] S. Hawking. *The Universe in a Nutshell (Bantam Books, New York 2001)*, 2016 dtv Verlagsgesellschaft mbH & Co. KG, München.

[4] K-H. Spatschek. *Astrophysik, eine Einführung in Theorie und Grundlagen*, (German) [*Astrophysics, an introduction to theory and basics*]. Spektrum Akademischer Verlag Heidelberg, 2nd edition, 2018.

[5] I.N. Bronstein, K.A. Semendjajew. Taschenbuch der Mathematik, (German) [*A Mathematics Paperback*]. Thun, Frankfurt am Main, 21st edition, 1984.

[6] W. Jung. *Physik*. Fischer Kolleg 3, 1973.

[7] R. Gross. *Physik III Optik und Quantenphänomene*, Walther-Meissner-Institut Bayerische Akademie der Wissenschaften, März 2013.

[8] W. Zinth. *Optik*, Ludwig-Maximilians-Universität München, Januar 2019.

[9] M. Bartelmann, M. Maturi. *Weak gravitational lensing*, Scholarpedia, 12(1):32440, 2017.

[10] A. Einstein. *Über den Einfluß der Schwerkraft auf die Ausbreitung des Lichtes.* (German) [*On the influence of gravity on the spread of light*]. Annalen der Physik, 322(10):898–908, 1911.

[11] M. Camenzind. *100 years of light diffraction*, Heidelberg University, 2019.

[12] M. Meneghetti. *Introduction to Gravitational Lensing*, Heidelberg University, 2018

[13] J. Pritchard. *Gravitational lensing*, Harvard University, 2017

[14] P. Schneider. *Extragalactic Astronomy and Cosmology*, Springer, 2006.

[15] Schneider, Ehlers, Falco. *Gravitational lenses*, Springer, 1999.

[16] C. Wei , G. Li, X. Kang, Y. Luo, Q. Xia, P. Wang, X. Yang, H. Wang, Y. Jing, H. Mo, W. Lin, Y. Wang, S. Li, Y. Lu, Y. Zhang, S. H. Lim, D. Tweed, W. Cui. *Full-sky Ray-tracing Simulation of Weak Lensing Using ELUCID Simulations: Exploring Galaxy Intrinsic Alignment and Cosmic Shear Correlations*, The Astrophysical Journal, 853:25 (18pp), 2018 January 20.

[17] K. Kuijken, *The Basics of Lensing: a unique tool for cosmology*, ASP Conference Series, Vol. xxx, 2003.

[18] J. Hartlap. *Ray-Tracing Simulations of Weak Gravitational Lensing*, Rheinischen Friedrich-Wilhelms-Universität Bonn, 2009.

[19] A. Udalski, M. Kubiak. *The optical gravitational lensing experiment: The discovery of three further microlensing events in the direction of the galactic bulge*, The Astrophysical Journal, 426:L69-L72, May 1994.

[20] S. Käs. *Multiparameter Analysis of the Belle II Pixeldetector's Data*, 2019.

[21] K. Dort. *Search for Highly Ionizing Particles with the Pixel Detector in the Belle 2 Experiment*, 2019.

[22] P. A. R. Ade *et al., Planck 2015 results. XIII. Cosmological parameters*, 2015.

[23] E. W. Kolb and M. S. Turner. *The Early Universe*, Front. Phys., vol. 69, pp. 1–547, 1990.

[24] F. Zwicky. *On the Masses of Nebulae and of Clusters of Nebulae*, Astrophys. J., vol. 86, p. 217, oct 1937.

[25] S. Dodelson. *Modern cosmology*, San Diego, CA: Academic Press, 2003.

[26] D. Clowe, M. Bradac, A. H. Gonzalez, M. Markevitch, S. W. Randall, C. Jones, and D. Zarit- sky. *A direct empirical proof of the existence of dark matter*, Astrophys. J., vol. 648, pp. L109–L113, 2006.

[27] G. W. Angus, B. Famaey, and H. Zhao. *Can MOND take a bullet? Analytical comparisons of three versions of MOND beyond spherical symmetry*, Mon. Not. Roy. Astron. Soc., vol. 371, p. 138, 2006.

[28] M. Deru, A. Ndiaye. *Deep Learning mit TensorFlow, Keras und TensorFlow.js*. Rheinwerk Verlag, Bonn 2019.

[29] A. Géron. *Hands-On Machine Learning with Scikit-Learn, Keras & TensorFlow*, O'Reilly 2019, Kiwisoft S.A.S.

[30] I. Goodfellow, Y. Bengio, A. Courville. *Deep Learning*, Massachusetts Institute of Technology, 2016.

[31] J. Brownlee. *Deep Learning for Computer Vision*, 2020.

[32] S. Hijazi, R. Kumar, C. Rowen. *Using Convolutional Neural Networks for Image Recognition*, IP Group, Cadence, 2019.

[33] A. Del Popolo, and P. Kroupa. *Density profiles of dark matter haloes on galactic and cluster scales*, A&A 502, 733–747, 2009.

[34] S. M. K. Alam, J. S. Bullock & D. H. Weinberg. *DARK MATTER PROPERTIES AND HALO CENTRAL DENSITIES*, Department of Astronomy, The Ohio State University, Columbus, 2005.

[35] Bayesianbiologist
`https://bayesianbiologist.com/2012/11/24/simulating-weak-gravitational-lensing/`

[36] M. D. Schneider, K. Y. Ng, W. A. Dawson, P. J. Marshall, J. Meyers, D. J. Bard. *PROBABILISTIC COSMOLOGICAL MASS MAPPING FROM WEAK LENSING SHEAR*, Schneider et al. 2016.

[37] D. Nelson,J. Annis. *Weak Lensing Mass Measurements of Nearby Galaxy Clusters*, University of California, Berkeley, 2008.

[38] B. Moore, S. Ghigna, F. Governato. *DARK MATTER SUBSTRUCTURE WITHIN GALACTIC HALOS*, The Astrophysical Journal, 524:L19–L22, 1999.

[39] E. Retana-Montenegro, E. Van Hese, G. Gentile, M. Baes, F. Frutos-Alfaro. *Analytical properties of Einasto dark matter haloes*, A&A 540, A70, 2012.

[40] S. W. Allen, A. E. Evrard, A. B. Mantz. *Cosmological Parameters from Observations of Galaxy Clusters*, Annual Review of Astronomy and Astrophysics, 2011.

[41] B. McBryan. *Living in a Low Density Black Hole, Non-Expanding Universe – Perhaps a Reflecting Universe*, Cornell University, 2013.

[42] P. Salucci, F. Nesti, G. Gentile, C. F. Martins. *The dark matter density at the Sun's location*, A&A 523, A83, 2010.

[43] A. Zitrin, M. Meneghetti, K. Umetsu, T. Broadhurst, M. Bartelmann, R. Bouwens, L. Bradley, M. Carrasco, D. Coe, H. Ford, D. Kelson, A. M. Koekemoer, E. Medezinski, J. Moustakas, L. A. Moustakas, M. Nonino, M. Postman, P. Rosati, G. Seidel, S. Seitz, I. Sendra, X. Shu, J. Vega, W. Zheng. *CLASH: The enhanced lensing efficiency of the highly elongated merging cluster MACS J0416.1-2403*, Astrophysical Journal, November, 2012.