A Deep Learning Framework for the Detection of Baryon Acoustic Oscillations in the Sub-millimetre Galaxy Population

by

Analuz Silva-Silverio

Doctoral Advisors:

Dr. María del Pilar Gómez-Gil, INAOE
Dr. David Omar Sánchez-Argüelles, INAOE

Technical Report No. CCC-23-005
January, 2023

Instituto Nacional de Astrofísica, Óptica y Electrónica

Luis Enrique Erro 1
Sta. Ma. Tonantzintla,
72840, Puebla, México.

A Deep Learning Framework for the Detection of Baryon Acoustic Oscillations in the Sub-millimetre Galaxy Population

Abstract

Deep learning (DL) methods now can solve high-dimensional learning tasks, such as studies required by Cosmology, which aims to describe the physical processes that conducted the evolution and contents of the entire Universe, and nowadays it has encountered a major challenge in generation, processing and analysis of data. Today, observations and theoretical models are moving along several front lines to solve inconsistencies on the current cosmological concordance model known as Lambda Cold Dark Matter (ΛCDM). Indeed, our research aims to find an improved way to calculate the Baryon Acoustic Oscillations (BAO) signal, in this case over the distribution of Sub-millimetre Galaxies (SMG), whose peak density is located at redshift $z = 2 - 3$, proposing it as a new cosmological tracer.

In order to overcome the current shortness in the number of objects available by single-dish observations, we propose as part of this framework a new conditional generative adversarial network (cGAN) architecture able to handle several regression labels as input conditions. This type of data generation is a current research opportunity in the state-of-the-art (SOTA), in computational Cosmology. Recently, deep generative models have been proposed for the generation of synthetic cosmological data, being considered as an accelerated process to obtain new simulations. Our new model should be able to generate reliable synthetic galactic population maps of different distributions, depending upon a given set of cosmological parameters as conditions.

To complete the framework, we propose a new architecture based on Capsule Networks (CapsNet) combining the constructive procedure of Geometric Deep Learning (GDL). Our model will take advantage of the ability of CapsNet to store pose information such as position, rotation, scale, etc., with their intrinsic relationships in a high dimensional vector space and the GDL constructive procedure is able to incorporate prior physical knowledge into the architecture. We hypothesise that this model will be a competitive image feature extractor, compared to other SOTA DL models. Also, our results will be compared to those derived from ΛCDM.

We expect that our framework will obtain an improved BAO measurement, optimised for large-scale structure (LSS) observations obtained by the TolTEC camera commissioned on the Large Millimetre Telescope (LMT).

Keywords— *Deep Learning, Cosmology, Adaptive Regression, Deep Conditional Generative Models, Baryon Acoustic Oscillations, Large Scale Structure.*

Contents

1 Introduction

Over the past decade the world has witnessed an experimental revolution epitomised by deep learning (DL) methods; previously unreachable high-dimensional learning tasks in computer vision, pattern recognition, segmentation, estimation, etc., are currently viable, for some specific applications and data distributions, within a suitable computational scale [1]. It is well known that DL has a wide range of applications in various sciences and engineering fields, including cosmology[1], the area of interest in our research.

Among other things, cosmology analyses the distribution and properties of sets of astronomical objects, called cosmological tracers, such as type-Ia supernovae (SNIa), cosmic microwave background (CMB) anisotropies or the large scale structure (LSS) distribution [3]; aiming to recover the properties of the Cosmos on its largest scales. The process of cosmological model discrimination, which consist of a statistical comparison of the theoretical models predictions (proven using simulations), with the actual matter distribution (drawn by the aforementioned tracers), requires a careful control over the biases and uncertainties derived from observations.

The current cosmological concordance model, known as Lambda Cold Dark Matter (ΛCDM), is supported by an ample evidence derived from the last few decades of astronomical observations by state-of-the-art (SOTA) instrumentation. Despite its successes, ΛCDM is far from being a complete description of the Cosmos, facing significant challenges; one of these is that the physical properties of most of its components are unknown, and therefore described as "dark" components. Other challenge is the existence of statistical disagreements (tensions) among the values of the cosmological parameters derived from tracers in the near universe (e.g. galaxy clusters, SNIa) and the far universe (CMB anisotropies) [4], prompting researchers to explore theoretical models and observations on several fronts.

The development of state-of-the-art (SOTA) detectors installed in large collecting area telescopes has been motivated by the challenges mentioned above [5]. Traditionally, these studies are carried out in the optical/infrared spectra, such as the Sloan Digital Sky Survey (SDSS) [6], the extended Baryon Oscillation Spectroscopic Survey (e-BOSS) [7], or the Dark Energy Spectroscopic Instrument (DESI) [8]). However, this thesis aims to explore the potential applications of the population of submillimetre galaxies (SMG) as a cosmological tracer, leveraging observations of a new detector called TolTEC [9], which is currently being commissioned on the Large Millimeter Telescope (LMT) (Spanish: Gran Telescopio Milimétrico, or GTM), the world's largest single-aperture telescope in its frequency range, located in Sierra Negra, Puebla [10].

The design of algorithms capable of obtaining novel information of the Cosmos through the analysis of data coming from observations, is another complement to overcome the aforementioned challenges. An important example of this is the measurement of the baryon acoustic oscillations (BAO) signal (see Section 2.1.2), a property of the LSS distribution that can be obtained as a characteristic scale that emerges from the early Universe ($\simeq 300\,\mathrm{Myr}$ after the Big Bang), whose effect left its imprint as a typical separation within the distribution of galaxies (an increment in the probability near $\simeq 480\,\mathrm{Mlyr}$).

BAO have been measured in the galaxies distribution through two-point correlations and the power spectrum of matter fluctuations as a function of scale. Many studies rely on observations over a wide area (typically several tens to hundreds of sq. deg.), where the amount and the spatial

[1]Cosmology is the study of the Universe regarded as a whole, so called the Cosmos, with the aim of explaining its origin and evolution, through a description of its underlying physical processes governing it [2].

distribution of the cosmological tracers are strongly coupled to a subset of cosmological parameters. As the BAO scale is a 'standard ruler', it has a constant value once the effect of the accelerated expansion of the Universe is subtracted. However, detecting BAO is a complex problem, since the signal is diluted by the non-linear evolution of the galaxy distribution. In the particular case of searching this signal in the SMGs distribution, the fact that it represents a limited subset of all galaxies requires overcoming the current shortage of available data.

A common way to mitigate this is to generate simulations following theoretical models (known as cosmologies), based on our knowledge on physics, providing a wider sample of what the Universe could look like. In this regard, N-body techniques[2] simulate the evolution of the Universe over cosmic time. However, such simulations' runtime is very slow and computationally expensive, in magnitudes around 10^7 CPU hours [12]. Examples of programming code for cosmological N-body simulations are L-PICOLA [13] and GADGET [14], and some simulations built this way are Bolshoi [15], Cosmo-π [16] or Illustris [17].

Deep generative models have been recently proposed for the simulation of cosmological data. It is considered an accelerated process to obtain synthetic data of cosmological tracers. However, these models present some drawbacks; following in particular the approach of generative adversarial networks (GANs), most works produce data limited to fixed values of cosmological parameters (namely "continuous conditions" or "regression labels"), which limits the applicability and re-use of these models [18]. Therefore, in this research, as part of our new framework, we propose a novel GAN model capable of using more than one continuous condition to generate galactic distribution maps, based on the conditioning cosmological parameters of matter density (Ω_m) and matter clustering strength (σ_8) (see Section 2.1.1), as they strongly impact on the formation of structure along the cosmic time.

To obtain the BAO signal from the SMGs distribution, this thesis proposes to complete our framework by the inclusion of a new Capsule Networks (CapsNet) model as an estimator [19]. Today convolutional neural networks (CNN) are the SOTA estimator models for many kinds of tasks, however, they present problems related to the lose of spatial information cause by the use of pooling layers. CapsNet have been proposed as a way to overcome these drawback. A capsule can stores mainly spatial information about positional attributes of objects in a high dimensional vector space, representing in each dimension a feature of the pattern about to be identified [19]. It is contemplated for the design of this model the incorporation of the geometric deep learning (GDL) constructive procedure, which provides a way to integrate prior physical knowledge into neural architectures [1]. GDL considers that even tough the learning of generic functions in high dimensions is a hard estimation problem, most common tasks come with important predefined regularities from the underlying dimensionality and the structure of the physical world [20]. This way, a combination of CapsNets elements and the GDL constructive procedure would improve features' extraction and pattern recognition from images, working in this case as an estimator to obtain the BAO signal.

Summarising, this thesis aims to build a new framework capable of generating synthetic galactic maps corresponding to the expected observations of TolTEC and recovering the BAO signal from the TolTEC maps of SMG. To overcome the limited number of objects in millimetre wavelength observations, we propose the development of a conditional GAN that supports more than one con-

[2]A N-body technique represents the distribution of matter in a 3D space, by trillions of particles, evolving their positions and velocities over cosmic time in small time intervals from the early Universe (when mass distribution is described by a Gaussian random field), to later times (when it becomes highly non-Gaussian due to several physical processes, such as gravitational instability growth, star formation, etc) [11].

tinuous condition. This will allow to generate synthetic maps of the galaxy distribution and to contrast them with real observations of TolTEC. The framework is completed with the recognition of the scale of BAO, which will be achieved through the design of a new CapsNet model incorporating the constructive procedure of GDL. This model will be trained using a combination of the synthetic data and real observations.

The main computational contributions expected from this work are in the components of the proposed framework. These include a new continuous conditioned generative DL model capable of managing more than one condition, and a novel DL architecture that combines properties of CapsNet with GDL to extract image features, being a competitive extractor against SOTA extractors working as estimators. On the other hand, the contribution to the field of cosmology is the increased understanding of the way SMG maps can be exploited, which potentially could give tools in finding the required adjustments to the cosmological concordance model.

This document presents in section 2 the basic concepts required to understand the problem and the approach concepts to be considered. Section 3 shows the related work, followed by section 4, where the research proposal is presented. Section 5 presents the methodology to achieve the objectives of this project. Section 6 presents the preliminary results of the experiments carried out so far, and section 7 comments on some remarks and perspectives of this research.

2 Basic Concepts

2.1 Cosmology

Cosmology studies the Universe as a whole, so called Cosmos. Etymologically, the word 'cosmos' comes from the Greek word 'kosmos', which means harmony or order. Cosmologists deals with fundamental questions concerning the human condition, such as 'What is the Universe made of? Is it finite or infinite in spatial extent? Did it have a beginning in the past? Will it come to an end in the future?'. Due to the immense complexity and richness of the Universe, condensing its story follows a process of ruthless simplification: the properties of an idealised, perfectly smooth model universe [2].

Today, cosmology is a mature experimental science. In recent years, theoretical research has focused on motivating observational projects and predicting the signals expected from these instruments. This way, the evolution of the Universe can now be explored through direct observations using large ground-based telescopes and space telescopes in outer space [5], some examples are the Large Millimeter Telescope (LMT), the Atacama Large Millimeter Array (ALMA), the Thirty Meter Telescope (TMT), the Extremely Large Telescope (ELT), the James Webb Space Telescope (JWST), among others. These instruments are committed to provide new images of the young Universe and the first galaxies.

Existing datasets include an image of the Universe when it was $\sim$0.4 million years old (the CMB aniostropies), and images of galaxies of the Universe when it was older than a billion years. The explanation challenge of the cosmological models is found between these two periods, the time when the Universe was dark: stars were not formed yet, and the CMB no longer traced the matter distribution. This is a pretty important period, as the primordial soup evolved into a rich variety of objects of the current Universe [5].

2.1.1 The Cosmological Concordance Model ΛCDM

A model of the Universe based on the cosmological principle[3] and an expanding universe[4] provides an accurate description of numerous observational facts. In General Relativity, the metric for a spatially homogeneous and isotropic space-time is the Friedmann-Lemaître-Robertson-Walker (FLRW) metric, a model used as a first approximation for the evolution of the Universe, which can be written in the form

$$ds^2 = c^2dt^2 - a^2(t)\left[\frac{dr^2}{1-kr^2} + r^2(d\theta^2 + sin^2\theta d\phi^2)\right], \tag{1}$$

where c is the speed of light, $a(t)$ is the cosmic scale factor (description of the expansion as a function of time t), and (r, θ, ϕ) are the comoving coordinates (a frame where observers remain at rest, but their physical separation increases with time in a factor given by $a(t)$). The constant k defines the geometry of space-time (A closed *spherical-like* universe is positive, a flat universe *Euclidean-space* is zero and an open *hyperbolic* universe is negative).

[3]The cosmological principle assumes that the matter and energy distribution is homogeneous and isotropic on the largest scales.

[4]A well-established property of the Universe is its constant expansion, first detected by Hubble while observing the recession speed of nearby galaxies and now confirmed by observing supernovae (see section 2.1.3).

A given observer sees a nearby object at a physical distance D moving away at a velocity of $v = H(t)D$, where the Hubble-Lemaître constant at time t is $H(t) = da(t)/dt$, the expansion parameter of the Universe. This measure of the speed of the expansion characterises fundamental features such as the age and size of the observable universe, and H_0 is its present value.

Photons emitted by a source at time t are observed at $t = 0$ with a wavelength redshift[5] of $z = 1/a(t) - 1$. This means that $a(t)^{-1} = (1 + z)$, where $a(t = 0) \equiv 1$ is set for convenience. Once the FLRW metric is defined, an analytic solution to the Einstein field equations can be reduced to the Friedmann equation

$$H^2(t) = \frac{8\pi G}{3}\rho - \frac{k}{a^2}, \tag{2}$$

relating the matter-energy content to the expansion rate of the Universe rate. Here G represents Newton's gravitational constant and ρ denotes the sum of the components contributing to the energy density of the Universe.

In order to describe the evolution of each component we need an equation of state that relates the density to the thermodynamic pressure $p = p(\rho)$. Therefore, the density ρ varies with $a(t)$ according to the thermodynamic relation $d(\rho c^2 r^3) = -pd(r^3)$.

Another important quantity is the critical density ρ_c

$$\rho_c(t) \equiv \frac{3H^2(t)}{8\pi G}, \tag{3}$$

which is defined as the density required in a matter-dominated universe to eventually halt its expansion. In the literature, it is a common practice to parameterise the total density ratio Ω to the critical density is

$$\Omega \equiv \frac{\rho}{\rho_c}. \tag{4}$$

The present contributions of all components to Ω are denoted as follows: Ω_m (matter, including cold dark matter Ω_{DM} and ordinary matter, so called baryonic matter Ω_b), Ω_Λ (vacuum energy - cosmological constant), Ω_r (radiation), and Ω_k (spatial curvature). Consequently, the Friedmann equation becomes

$$\frac{H(t)}{H_0} = \left[\frac{\Omega_m}{a^3} + \Omega_\Lambda + \frac{\Omega_r}{a^4} + \frac{\Omega_k}{a^2}\right], \tag{5}$$

where $H_0 = 100, h, \mathrm{km\,s^{-1}\,Mpc^{-1}}$ in terms of h (the dimensionless Hubble constant, commonly referred to as 'little h,' often used when expressing distances calculated from redshift z instead of using H (Croton, 2013)). The variable $\Omega_0 = \Omega_m + \Omega_\Lambda + \Omega_r$, and H_0 represent the present values of H and Ω, respectively. The overall normalisation of the power spectrum is specified in terms of σ_8 (the root-mean-square amplitude of mass fluctuations in spheres of radius $8h^{-1}\mathrm{Mpc}$), and the geometry term is:

[5]Cosmological redshift is a measure of distance using the formula $d \approx c/H_0 \times z$, where $z = 0$ represents the present universe. It occurs when the light from distant galaxies is stretched over time by the expanding space. As the light travels, the wavelengths are increased (reddening), which should not be confused with the Doppler effect caused by the motion of the light source itself [21].

$$\Omega_k \equiv -\frac{k}{H_0^2} = 1 - \Omega_0. \tag{6}$$

The study of the energy components of the Universe (see Fig. 1), and its properties can be done by cosmological tracers, which are the observations of astronomical objects related to the parameters mentioned through these theoretical models. Presently, recent observations have confined the standard set of cosmological parameters to a relatively narrow range. The values of the main parameters that best fit the WMAP data, the large-scale structure surveys of galaxies and supernovae are $\Omega_m = 0.28$, $\Omega_\Lambda = 0.72$, $\Omega_b = 0.046$, $\sigma_8 = 0.81$, and $h = 0.72$.

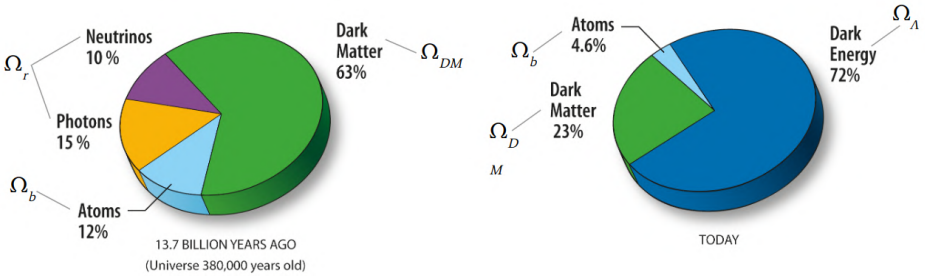

Figure 1: Energy density of components back then 13.7 billion years ago and today [22].

Today, it seems that the expanding Universe is dominated by a positive cosmological constant (Λ), characterised by a large density parameter Ω_Λ and cold dark matter (CDM), with a density parameter Ω_{DM}, being a ΛCDM cosmology with a Ω_k assumed to equal zero.

Current Challenges of ΛCDM

The Cosmological Constant (Λ) The cosmological constant Λ was introduced by Einstein in the equations in order to obtain a static Universe. However, it currently represents an unknown component that is thought to be driving the accelerated expansion of the Universe [23]. Originally, it was thought that the cosmological constant could be the vacuum energy; however this leads to three "problems":

First, the value of Λ is extremely small (approximately 122 orders of magnitude smaller than the theoretical expectations derived from quantum physics). If ρ_Λ were even slightly larger, the repulsive force would cause a rapid expansion, preventing the formation of gravitationally bound systems. This is known as the cosmological constant problem.

Second, the question is why the sum $\Omega_0 = \Omega_m + \Omega_\Lambda$ is precisely 1.0 today (within a margin of 0.3 percent), and why Λ has been fine-tuned to come to dominate the sum only in recent times. This is known as the cosmic coincidence problem.

Third, we only know that Λ energy distorts the geometry of the Universe as if it were matter with a strongly negative pressure. Since we have limited knowledge about it, we cannot be sure whether Λ is constant over time or if its equation of state is always $\omega = -1$.

The H_0 Tension Also known as the Hubble tension, revolves around the value of H_0. As an informative parameter, obtaining an accurate value is crucial for understanding the evolution of the

Universe [24], but measurements taken in both, the early and late universe, have revealed a statistical inconsistency in the derived values for H_0.

Using late-type tracers (see Section 2.1.3), such as the motions of galaxies, the value is $73\mathrm{km\,s^{-1}\,Mpc^{-1}}$; On the other hand, using early-type tracers (see Section 2.1.3), e.g., the CMB, the value is around $67\mathrm{km\,s^{-1}\,Mpc^{-1}}$, as seen in Fig. 2. The reason there is now so much literature about this tension stems from a new, more accurate surveys of the motions of nearby galaxies, which reduced the error bars around $73\mathrm{km\,s^{-1}\,Mpc^{-1}}$, surpassing the 5-sigma significance level.

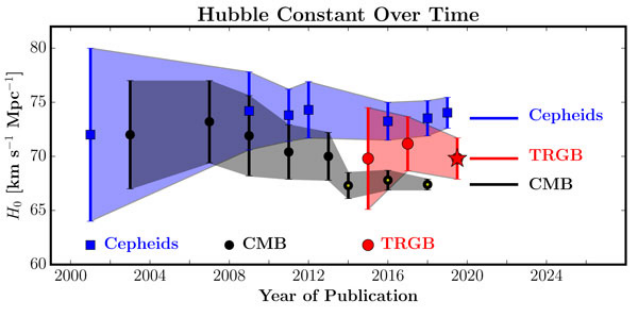

Figure 2: Hubble Constant over time. A plot of H_0 values as a function of time [24].

This tension implies that either the measurements of celestial objects in the universe today are inaccurate, or something about the current cosmological model is wrong, necessitating the introduction of new physics as one of the possibilities [24]. Cosmologists are hesitant to tinker with the ΛCDM model, as it provides remarkably accurate predictions for other features of the Universe, resulting in minimal consensus regarding which of these options is correct.

The σ_8 Tension A similar tension arises in the calculation of the σ_8 parameter [25]. It appears that the observational probes of the growth of density perturbations indicate a possible weakening of the force of gravity at low redshifts z. The evidence is at about a 2-3σ level and primarily comes from weak lensing data, which measure the parameter $S_8 = \sigma_8\sqrt{\Omega_{0m}/0.3}$ and redshift space distortion data that measure the growth rate multiplied by the amplitude of the linear power spectrum parameter $f\sigma_8(z)$. The measured $f\sigma_8$ seems to be lower than the prediction of the General Relativity Theory within the context of the ΛCDM model, as defined by the best fit parameter values from the Planck survey [25].

2.1.2 Origin of BAO

Thermal Beginning of the Universe The thermal history of the Universe is based on the extrapolation of the present knowledge of the Universe and particle physics back to the Planck epoch ($t \approx 10^{-43}\,\mathrm{s}$ (seconds) or $T \approx 10^{19}\,\mathrm{GeV}$ (gigaelectron-volts)), the moment at which quantum corrections to general relativity are thought to set in. At the earliest stages, the Universe was a hot plasma of relativistic particles (including quarks, leptons and Higgs bosons) [26]. After this, during the inflationary epoch occurred between $t \approx 10^{-35}\,\mathrm{s}$ and $t \approx 10^{-24}\,\mathrm{s}$, the metric that defines distance within space, underwent a sudden changed in scale, resulting in the early Universe expanding to at least 10^{78} times its previous volume at a speed much greater than speed of light, to become the visible Universe of today. As our local portion of the Universe was much more smaller, it shared

properties such as temperature, density, distribution or geometry, which remained extremely homogeneous after this expansion. This theory, first proposed by Alan Guth in 1979, solves the horizon problem, the flatness problem, the mono-poles and temperature [26].

In the following epoch, called nucleosynthesis (which follows when $t \approx 10^{-2}\,\mathrm{s}$ to $t \approx 10^{2}\,\mathrm{s}$ and $T \approx 10\,\mathrm{MeV}$ to $T \approx 0.1\,\mathrm{MeV}$), the Universe was dominated by a hot plasma, period known as the radiation dominated era. The high temperatures prevented atomic nuclei from holding onto electrons for long periods and most of the baryonic matter[6] was in a plasma state. When the matter density became equal to that of radiation, the matter dominated epoch began ($t \simeq 31000\,\mathrm{yr}$), implying the start of structure formation. Finally, at an age of $t \simeq 3.17\,\mathrm{Myr}$, the overall Universe temperature dropped enough with the expansion, allowing ions and electrons to combine and form the first neutral atoms. This capture process releases photons, enabling them to propagate freely through space and initiate the matter-radiation decoupling, thereby ending the epoch of near thermal equilibrium.

Cosmic Microwave Background The large population of photons that propagated freely through the Universe is detected today at wavelengths of a few millimetres and is known as the cosmic microwave background (CMB), which is the farthest observable light in the Universe [2].

Nowadays the analysis of the CMB anisotropies (see Fig. 3), is one of the pillars of precision cosmology. The first theoretical cosmological calculations predicting CMB emission were made in the 1940s, but they were confirmed until its accidental detection by Penzias and Wilson in 1964 [2]. The theoretical models also predict that CMB radiation must have a thermal black-body spectrum and inflation predicts that the CBM's mean temperature should exhibit arc-minute perturbations across the sky [28]. The Cosmic Background Explorer (COBE), a satellite experiment launched in 1990, verified these predictions while detecting an emission consistent with a black-body spectrum with a precision of a few hundreds of μK on angular scales larger than 7º. Various experiments have since extended the range, such as the Wilkinson Microwave Anisotropy Probe (WMAP) and more recently the space mission Planck have significantly improved precision and statistics at different angular scales [28].

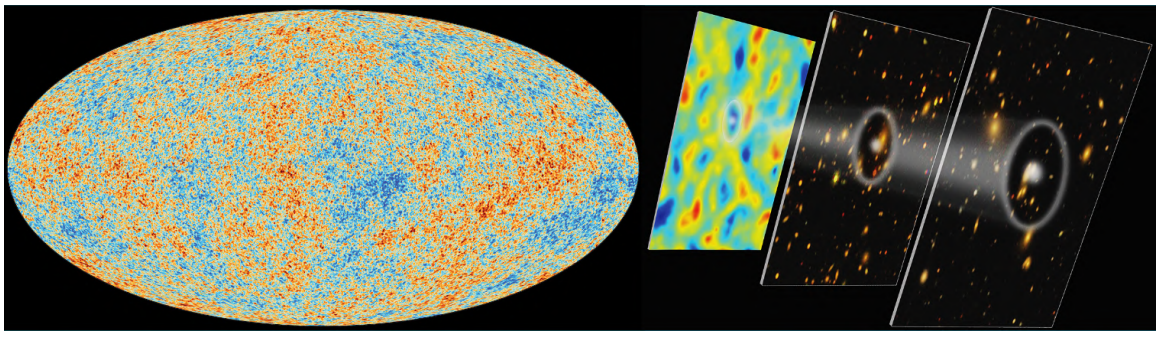

Figure 3: Left: CMB Anisotropies taken by Planck/ESA [29]. Right: A cartoon of the record of BAOs (white rings) in maps of galaxies. This record helps to retrace the history of the universe expansion [30].

[6] In astrophysics, the term "baryonic matter" refers to all objects made of normal atomic matter, essentially ignoring the presence of electrons which represent only 0.0005 of the mass [27].

Baryon Acoustic Oscillations During the mentioned nucleosynthesis epoch, the hot plasma was a dense almost uniform fluid with tiny fluctuations (of about one part in 100,000), forming slightly denser kernels of matter. These clumps attracted more surrounding matter through gravity, but particles could not stick together because of the high temperatures. The interplay between gravity's pulling force and the repelling effect resulted in pressure waves (sound waves), that propagated through the plasma, known as baryon acoustic oscillations (BAO). BAO *characteristic scale* is defined as the distance the pressure waves were able to travel at the matter-radiation equilibrium epoch [31]. When matter-radiation decoupling began, particles stopped repelling each other, waves ceased and the ripples were essentially frozen in place, becoming a part of the CMB anisotropies and carrying within them a bit more matter than the average with the Universe evolution (see Figure 1).

The BAO traces still linger, showing their imprint on the Cosmos (see Figure 4). During the Dark Ages, gravity became the dominant force, the clumps grew by the accretion of material, as this process was slightly more efficient along the density ripples and with a sufficiently great density drove to the formation of structure, i.e stars, agglutinated together to form galaxies and galaxies clusters[7]. The frozen ripples were stretched with universe expansion and the peculiar separation of BAO between galaxies also increased.

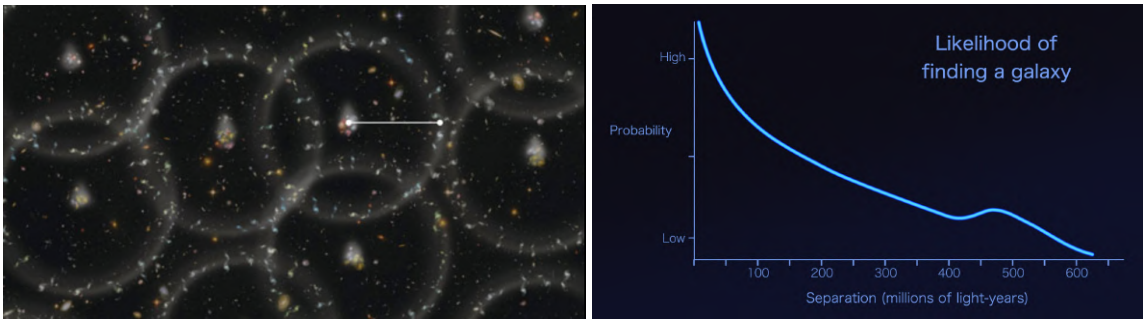

Figure 4: Left: A cartoon produced by the BOSS project showing the spheres of baryons around the matter clumps. Right: Probability of finding a galaxy pair as a function of the distance between them [31].

BAO scale is well defined by theoretical models, and can be used as a cosmic ruler[8] and through this derive the rate at which the Universe is expanding H_0, being essential to put constraints on σ_8, as well as on the Λ [32]. The small increment in the probability near $\simeq 480\,\mathrm{Mlyr}$ shows the influence of the BAO to the galaxy distribution (see Figure 2). This way, BAO is observable in the power spectrum of the CMB anisotropies (see Figure 3), and in the two-point correlation function of the galaxies distribution known today (see Figure 4).

2.1.3 Cosmological Tracers

Late Type Tracers Virtually, the only data we have about distant celestial objects is how bright they appear to telescopes. Therefore, there are certain celestial light sources such as some stars and supernovae (SNIa), so called "standard candles", with predictable luminosities to determine their distance from our location [33].

[7] Clusters are fully collapsed objects that have come to certain equilibrium, obeying the virial theorem [2].

[8] A cosmic ruler is a fixed size-distance that can be measured at different epochs, coupled to the evolution of the universe, decreasing with the look-back time [31]

Once the distances to these standard candles in the Universe are established, we obtain a "map" of the observable universe known as the "cosmic distance ladder". This allows us to plot the velocities of galaxies by using the spectral line redshift z against their distance from us and the gradient of this plot provides an estimate for the H_0 [33]. However, these calculations have historically been ridden with errors[9]. Some cosmologists argue that there are systematic errors resulting from incorrect measurements in the distance ladder, which means that the perception of "the H_0 tension" is not a fundamental disagreement between theory and observations [24].

Early Type Tracers The CMB radiation was key evidence in favour of the expanding universe theory over older, discarded cosmological theories. The phenomenology of the acoustic peaks in the temperature and polarisation of the power spectrum of CMB fluctuations is mainly described by four observables and the initial conditions [34]. These observables are the angular extents of the sound horizon $\ell_{\rm a} \equiv \pi D_*/s_*$, the particle horizon at matter-radiation equality $\ell_{\rm eq} \equiv k_{\rm eq} D_*$, the damping scale $\ell_{\rm d} \equiv k_{\rm d} D_*$, and the value of the baryon-photon momentum density ratio R_* (see Figure 5). The initial conditions determine the phase, or equivalently the location of the first peak in units of $\ell_{\rm a}$, and an overall tilt n in the power spectrum [34].

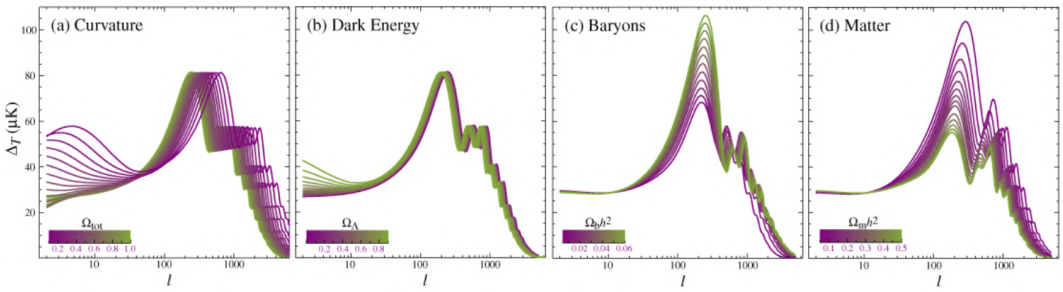

Figure 5: Power spectrum of CMB anisotropies for different sets of cosmological parameters [34].

The CMB can give an estimate of H_0 back at the time the CMB was formed, and then through simulate the Universe evolution from that point to obtain the expected value for the constant today, giving a theoretical prediction for the value of H_0 [33]. If the theoretical prediction is wrong, then there must be a problem with the cosmological model.

Large Scale Structure Tracers In cosmology, the so-called large-scale structure of the Universe (LSS) represents all the structures larger than individual galaxies, where the most important structures are the superclusters and voids. A supercluster contains one or more galaxy clusters (see section 2.1.2), embedded within it, which can be either flattened (roughly planar) or elongated (roughly linear) structures, while under-dense voids are roughly spherical in shape. The detection of clusters by mapping the Universe with massive galaxy surveys is a promising approach to provided critical cosmological information to reveal the physics governing the evolution of the Universe.

Last century has seen a great progress on galaxy surveys. Since the 1970s, larger galaxy surveys have been carried out given the technical advancements in multi-object spectrographs and

[9]The original estimate for the H_0calculated by Hubble was approximately $500 \mathrm{kms}^{-1}\mathrm{Mpc}^{-1}$, which is significantly larger than the current settled value of $73 \mathrm{km\,s^{-1}\,Mpc^{-1}}$.

charge coupled device (CCD)-based photometry [32], such as the SDSS project (based on the 2.5-meter Sloan Telescope), one of the most powerful and successful galaxy surveys along the last two decades, which has created the largest map of the Universe through the observations of various kinds of tracers (see Figure 6) [35].

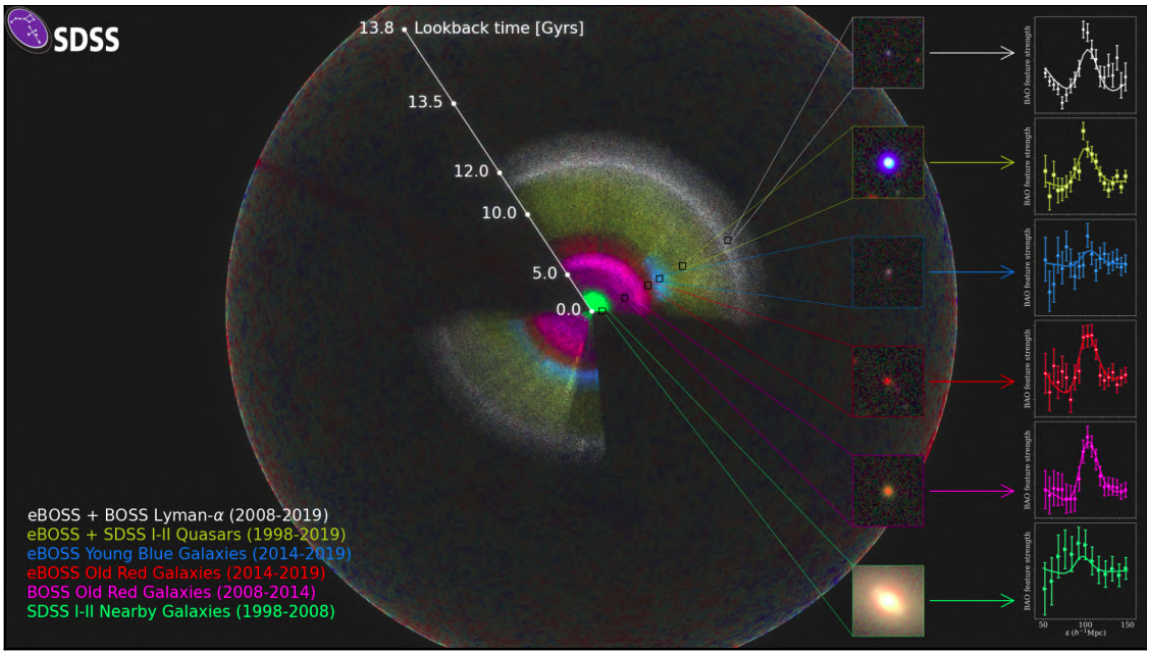

Figure 6: Mapping the Universe in three dimensions with SDSS (1998-2020) using different types of tracers, as illustrated using different colours. The outermost is the last scattering surface of the CMB. The six panels on the right are the BAO-peak measurements using different samples, as labelled in the legend in the lower-left corner [35].

Using the galaxies distribution, a map of the LSS can be traced by measuring the redshift z of a sample of galaxies and applying the Hubble relation $d = (c/H_0)z$ to compute their distances from our own Galaxy [2]. By plotting the redshift z as a function of angular position for galaxies in a long, narrow strip of the sky, a "slice of the Universe" can be mapped. Another measure obtained through the LSS tracers is BAO, which primarily encode key cosmological information, corresponding to specific clustering patterns of biased tracers, measured from nearly all available wide-field galaxy surveys. As a standard ruler, the BAO measurement complements other probes such as SNIa and CMB, with the advantage of not requiring calibration using a distance ladder measured in the nearby universe like a standard candle (e.g., SNIa) [36].

Sub-millimetre Region Tracers Until recently, millimetre/submillimetre experiments did not have the sensitivity to image both wide and deep, to distinguish among competing models of galaxy evolution. One tracer in the mm-wave regime is the Sunyaev-Zel'dovich Effect (SZE)[10] (see Figure 7, which has been use to trace the LSS via galaxy clusters [36]. Another potential tracer is the population of submillimeter galaxies (SMGs), which informs models of the formation of structure

[10]Sunyaev-Zel'dovich Effect occurs when photon from CMB traverse the atmosphere of galaxy clusters, gaining some energy. The net effect is a CBM energy deficit seen at certain wavelengths (long wavelengths) [36].

in the Universe [37], but a significant expansion of these surveys is required in order to know how is the distribution of these dust-obscured galaxies within the LSS.

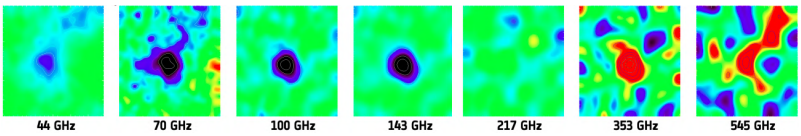

Figure 7: Sunyaev-Zel'dovich Effect. Multi-band observations of the galaxy cluster Abell 2319 [38].

2.1.4 Data Acquisition

The data acquisition system of the LMT that coordinates the telescope with its instruments to produce scientific data, is a new software developed following the experience obtained from the 14m FCRAO [10]. This new system was used successfully with previous testing instruments (the optical camera and holography system) and with the first two scientific instruments (AzTEC and RSR).

Currently commissioned at the LMT, TolTEC is a novel continuum, large format camera. It has the capability to map the sky in three wavelengths simultaneously ($\lambda = 1.1, 1.4, 2.1mm$) with an instantaneous field-of-view of $\simeq 4$ arcmin. The combination of a large number of detectors ($\sim$ 6500) and the sensitivity provided by Lumped Element Kinetic Inductace technology (LEKID) will allow to produce observations over large area of the millimetre sky with an unprecedented angular resolution (up to 5 arcsec at 1.1mm) Fig. 8 [9].

Wavelength	Beamsize (FWHM)	Number of Pixels	Number of Detectors	**Maximum** Mapping Speed Prediction
2.0mm	9.5 arcseconds	450	900	69 deg^2/mJy^2/hr
1.4mm	6.3 arcseconds	900	1800	20 deg^2/mJy^2/hr
1.1 mm	5 arcseconds	1800	3600	12 deg^2/mJy^2/hr

Figure 8: Key features of the camera TolTEC. Note: All detectors are polarisation-sensitive. Mapping speed values are for opacity of $\tau_{225GHz} = 0.1$ [9].

TolTEC will offer imaging of the sky at millimetre wavelengths through a series of Legacy Surveys that will result in publicly and freely available data sets. It is programmed to have the first data release of the first half of the surveys out in the summer of 2023 and the second data release approximately one year later. From the initial four surveys, two are of interest to our project:

- **100 sq. degree Large Scale Structure Survey.** It probes the relationships between the spatial distribution of SMG, so called star forming galaxies and LSS, and provides a detailed view of clusters and their substructure via the SZ effect.
- **1 sq. degree confusion limited Ultra-Deep Galaxy Survey.** This survey is designed to provide, for the first time, an unbiased sample of typical luminous infrared galaxies (LIRGs), ultraluminous infrared galaxies (ULIRGs) ranging from less to more luminous and virulent star-forming galaxies, across cosmic time from redshifts (z) 2 to 10. Over 55,000 such galaxies are expected to be detected in this survey in addition to over 200 strongly lensed and

extremely luminous galaxy systems. The main goal is to provide a complete history of the stellar mass build-up across the first 3 Gyr of from the Universe origin.

Dataset formats The raw data are usually time series, which contain information about the brightness of astronomical objects and the movement of the telescope. The data is processed to remove "contaminating" signals (e.g. cosmic rays, atmospheric and instrumental noise) and projected onto a two-dimensional grid of arbitrary size, which is known as a map [39]. In the case of TolTEC, the input set consists of three maps (one for each wavelength) with a pixel size smaller than 1 arcsec. The distribution format of these maps are both of type NetCDF (Network Common Data Form), as well as in the most widely used in the astronomical community, the FITS (Flexible Image Transport System) format.

NetCDF (Network Common Data Form), is a set of interfaces for array-oriented data access and a freely distributed collection of data access libraries for C, Fortran, C++, Java, and other languages. The NetCDF libraries support a machine-independent format for representing scientific data [40].

FITS (Flexible Image Transport System) is the most widely format used in astronomy for transporting, analysing, and archiving scientific data files [41]. It is primarily designed to store scientific datasets consisting of multidimensional arrays (images) and 2-dimensional tables, organised into rows and columns of information. A FITS file is comprised of segments called Header/Data Units (HDUs) [41].

2.2 Deep Learning Strategies

2.2.1 Basic Concepts

Machine Learning In computer science, learning can be defined as follows: *"A computer program is said to learn from experience E with respect to some class of task T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E"* [42]. Then, with this definition in mind, a machine learning (ML) algorithm is an algorithm that is able to learn from data.

ML addresses the question of how to build computer programs that improve their performance at some task through experience [42]. ML tasks are usually described in terms of how the ML system should process an example [43]. Many kinds of tasks can be solved with ML. The tasks of our interest are:

- **Density estimation or probability mass function estimation.** In the density estimation problem, an algorithm learns a function $p_{model} : \mathbb{R}^n \rightarrow \mathbb{R}$, where p_{model} is a probability density function (if $\boldsymbol{x}$ is continuous) or a probability mass function (if $\boldsymbol{x}$ is discrete) on the space that the examples were drawn from. To do so, the algorithm needs to learn the data structure it is seeing, e.g., if we have performed density estimation to obtain a probability distribution $p(\boldsymbol{x})$, we can use that distribution to solve the missing value imputation task. If a value x_i is missing and all of the other values, denoted $\boldsymbol{x_{-i}}$, are given, then we know the distribution over it is given by $p(x_i|\boldsymbol{x_{-i}})$ [43].

- **Synthesis and sampling.** Here, the algorithm has to generate similar examples to those in the training dataset. Synthesis and sampling via ML can be useful for media applications where it can be expensive to generate large volumes of content. In some cases, we want the sampling

or synthesis procedure to generate some specific kind of output given the input. There are structured output task, with the added qualification that there is no single correct output for each input. In order to produce an output that seems more natural and realistic, it is desired to have a large amount of variation in the input [43].

- **Imputation of missing values.** In this task, a ML algorithm receives a new example $\boldsymbol{x} \in \mathbb{R}^n$, with some entries x_i of $\boldsymbol{x}$ missed, and it must obtain a prediction of the values of the missing entries [43].

- **Regression.** For this task, the algorithm has to predict a numerical value given some input. To solve this, the algorithm is asked to output a function $f : \mathbb{R}^n \to \mathbb{R}$. Even though it is similar to classification, the significant difference is the format of output [43].

Supervised and Unsupervised Learning A subcategory of ML called Supervised learning, also known as supervised ML, is defined by its use of labelled datasets that yields the desired output to train algorithms capable of classifying data or predicting outcomes accurately. This training dataset includes inputs and correct outputs, which allow the model to learn over time. Among the various algorithms and computations techniques used in supervised ML processes, the most commonly used are artificial neural networks (ANN), naive bayes, linear regression, logistical regression, support vector machines (SVM), k-nearest neighbour (KNN) and random forest [44].

Unsupervised learning is a less clearly defined branch of ML, as there are many different unsupervised learning algorithms pursuing many different goals. Broadly speaking, the goal of unsupervised learning is to learn something useful by examining a dataset containing unlabelled input examples. Clustering and dimensionality reduction are common examples of unsupervised learning [45].

2.2.2 Deep Learning

The ANNs process a dataset by imitating the connectivity of brain neurons through layers of nodes. Each node is made up of inputs, weights, a bias (or threshold), and an output, emulating a neuron. If that output value exceeds a given threshold, it activates the node, passing data to the next layer in the network. ANNs learn this mapping function through supervised learning, adjusting based on the loss function through the process of gradient descent, among others. When the cost function is at or near zero, model's accuracy is supposed to yield the correct answer [44].

According to [43] learning networks have passed through three stages that they refer as cybernetics (around 1940-1960), connectionism (around 1980-1990) and deep learning (DL) (since 2010).

During the cybernetics stage, learning algorithms were intended to be computational models of biological learning, the biological neurons, giving birth to the name Artificial Neural Networks (ANN). Connectionism arose in the context of cognitive science, an interdisciplinary approach for understanding the mind, where the central idea is that a large number of simple computational units can achieve intelligent behaviour when networked together. Goodfellow and followers assure that DL stage corresponds to a more general principle of learning multiple levels of composition, which can be applied in ML frameworks that are not necessarily neural-inspired.

Today, DL is a wide subbranch of ML; we are able to train deeper neural networks than before, and to focus attention on the theoretical importance of depth [46]–[49].

The main recent progresses in Deep Learning Networks (DLNs) have been in Supervised Learning systems, nowadays widely used in a variety of applications; some of these include object detection and image recognition [44]. In the other hand, some applications imply additionally the generation of patterns as part of the output, so-called Generative Models.

2.2.3 Geometric Deep Learning

Geometric Deep Learning (GDL) is a niche within deep learning (DL) that aims to provide a common mathematical framework for studying the neural architectures that are currently used ubiquitously [1]. Additionally, it seeks to offer a constructive procedure for incorporating knowledge based on principles.

The General GDL Model The implementation of fundamental principles of symmetry (invariant functions, equivariant functions, prior scale separation, etc.) in the form of inductive biases, leads to some of the most popular architectures such as CNNs, Graph Neural Networks (GNN), Transformers, and Long Short-Term Memory (see Figure 9). This parallels how all of physics can be derived from their respective symmetries. We can refer to this as the DL-physical symmetry correspondence.

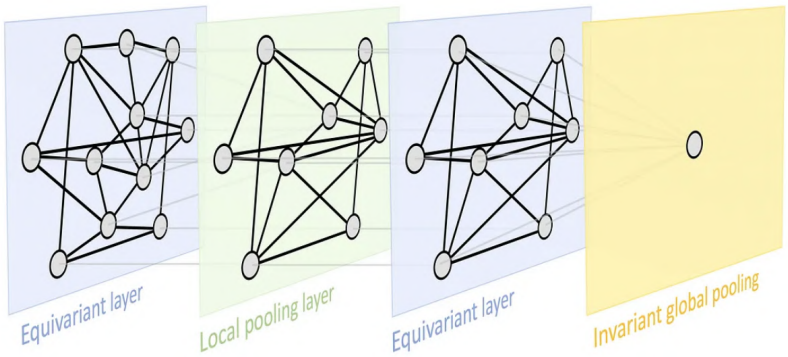

Figure 9: Geometric Deep Learning blueprint [50].

As this framework extensively operates within the input domain structure (geometric domains), it can be applied to any type of geometric data. The '5G' of GDL consists of Grids, Groups (homogeneous spaces with global symmetries), Graphs (including sets as a specific case), and Manifolds (see Figure 10). These elements demonstrate geometric priors through global isometry invariance, which can be expressed using Geodesics, as well as local Gauge symmetries [50].

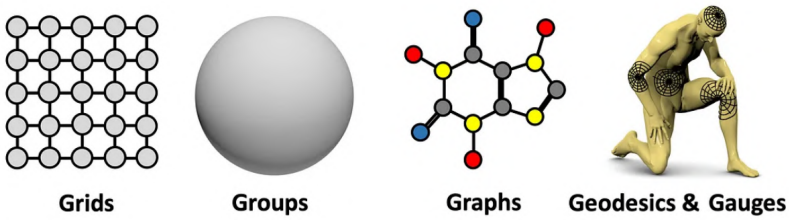

Figure 10: The '5G' of GDL [50].

GDL defines an emerging field of research that employs deep learning on non-euclidean data [1]. The advantage of interpreting the mesh in a non-euclidean way is that the geodesic distance is more meaningful for tasks performed on it. Using non-euclidean data, the shortest valid path between two points is not the Euclidean distance between them. The euclidean distance between points A and B is the length of the shortest straight path between them. A geodesic distance is a generalisation of the concept of the shortest path for higher dimensions, while the geodesic distance definition for graphs is usually the shortest path between nodes.

GDL Networks A GDL network, a kind of a graph neural network (GNN), is a neural model that capture the dependence of graphs via message passing between the nodes of graphs [51]. Many problems in physics for example, involve data in the form of unordered sets with rich relations and interactions, which can be naturally represented as graphs. This way, GDL constructive procedure provides a way to incorporate prior physical knowledge into neural architectures, hence, GDL provides a way to create new GNN model architectures.

A GNN is an tunable transformation on all attributes of a graph (nodes, edges, global-context), as seen in the example shown in Fig. 11 that preserves graph symmetries, which is known as permutation invariance.

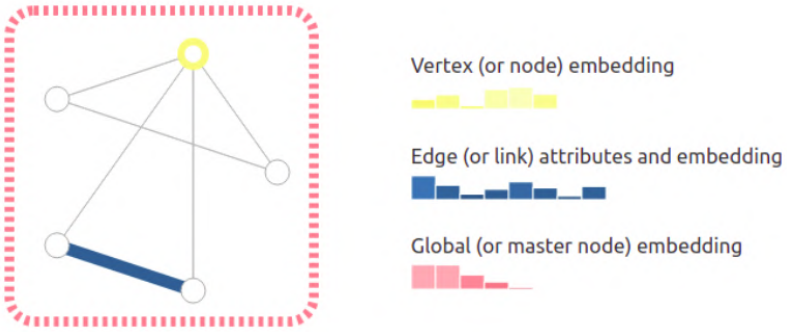

Figure 11: Information in the form of scalars or embeddings may be stored at each graph node (left) or edge (right). Here color bars represent the embeddings elements[52].

GNNs adopt a “graph-in, graph-out” architecture as showed in Fig. 12, meaning that these model types accept a graph as input, with information loaded into its nodes, edges and global-context, and progressively transform these embeddings, without changing the connectivity of the input graph.

2.2.4 Capsule Networks

Capsule networks (CapsNets) are a new architecture in neural networks, presenting an advanced approach to previous neural network designs, particularly for computer vision tasks [53].

Is is known that the brain is organised into modules, that may be considered capsules. With this in mind, CapsNets design incorporates dynamic routing algorithms to estimate features of objects like pose (position, size, orientation, deformation, velocity, albedo, hue, texture, and so on) [19].

Capsules Unlike normal neurons, capsules perform computations on their inputs and then ”encapsulate” the results into a small vector of highly informative outputs. A capsule could be considered

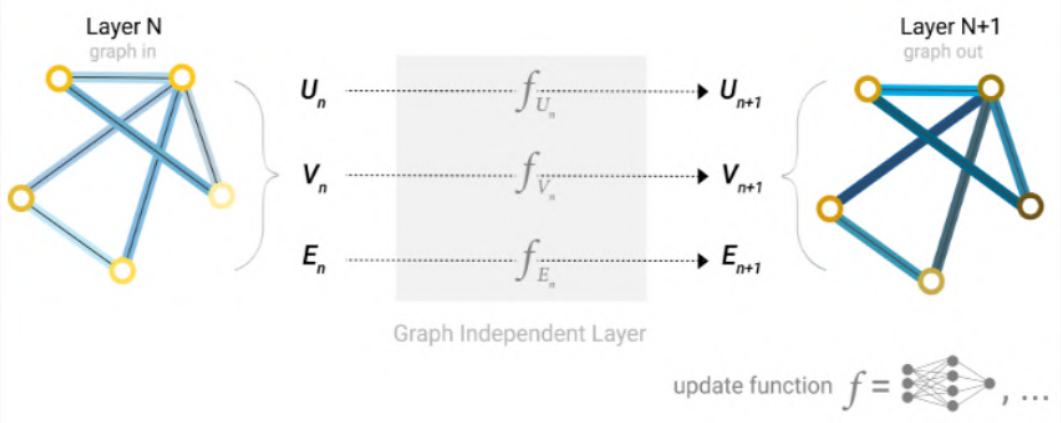

Figure 12: A single layer of a simple GNN. A graph is the input, and each component (V,E,U) gets updated by a MLP to produce a new graph. Each function subscript indicates a separate function for a different graph attribute at the n-th layer of GNN model [52].

a replacement or substitute for an average artificial neuron; whereas an artificial neuron deals with scalars, a capsule deals with vectors (see figure 13).

		capsule	vs.	traditional neuron
Input from low-level neuron/capsule		vector(u_i)		scalar(x_i)
Operation	Affine Transformation	$\hat{u}_{j\|i} = W_{ij} u_i$ (Eq. 2)		–
	Weighting	$s_j = \sum_i c_{ij} \hat{u}_{j\|i}$ (Eq. 2)		$a_j = \sum_{i=1}^{3} W_i x_i + b$
	Sum			
	Non-linearity activation fun	$v_j = \frac{\|s_j\|^2}{1+\|s_j\|^2} \frac{s_j}{\|s_j\|}$ (Eq. 1)		$h_{w,b}(x) = f(a_j)$
output		vector(v_i)		scalar(h)

Figure 13: Capsule vs Artificial Neuron [54].

For instance, we could summarise the steps taken by an artificial neuron as follows:

- 1. Multiply the input scalars with the weighted connections between the neurons.
- 2. Compute the weighted sum of the input scalars.
- 3. Apply an activation function (scalar nonlinearity) to get the output.

A capsule, on the other hand, goes through several steps in addition to those listed above, in order to achieve an affine transformation of the input, preserving co-linearity and ratio of distances [19]. Here, the process is as follows:

- 1. Multiply the input vectors by the weight matrices- This encodes spatial relationships between low-level features and high-level features.
- 2. Multiply the result by the weights.
- 3. Compute the weighted sum of the input vectors.
- 4. Apply an activation function (vector non-linearity) to get the output.

2.2.5 Deep Generative Models

The deep generative models may be seen as producing representations of probability distributions over multiple variables [45]. The main challenge with generative models is the training of examples x which are drawn from an unknown distribution $p_{data}(x)$.

The goal of a generative modelling algorithm is to learn a $p_{model}(x)$ that approximates $p_{data}(x)$ as closely as possible, such that the model generated data distribution is $p_{model}(x) = p_{data}(x)$ [55].

Some of these generative models let the probability distribution function to be evaluated explicitly, others instead, support operations that implicitly require knowledge of it. There are structured probabilistic models described in terms of graphs and factors[11], through the language of graphical models, while others cannot easily be described in terms of factors, but can represent probability distributions [56].

In order to reduce both the amount of human supervision and the number of examples required for learning, many researchers today study unsupervised learning, often using generative models [45]. This way, as generative models are assembled, they can be consider both supervised and unsupervised, depending on the components of its particular architecture.

Generative models that produce novel representative samples from high-dimensional data distributions are increasingly becoming popular in various fields. There are many different deep learning approaches to generative models. According to [11], the most popular ones are Variational Auto-Encoders (VAE), Autoregressive models, and Generative Adversarial Networks (GAN).

GANs are among the most successful generative models as they have the ability to generate highly realistic resolution images. GANs have been successfully applied to a wide variety of tasks, mostly in research settings. However, GAN still show unique challenges and research opportunities, among other reasons, due to the fact that they are based on game theory, while most approaches to generative modelling are based on optimisation [45].

Generative Adversarial Networks As mentioned, a generative model aims to construct a density estimator of a given dataset. The generative adversarial network (GAN) framework, first proposed by [56], constructs an implicit density estimator that can be efficiently sampled to generate samples of $\mathbb{P}_g$.

GAN establishes a mini-max game between two neural networks: a generator and a discriminator. The generator's training aims to generate samples undifferentiated from training data for a

[11] A factor in math is any number that can be multiplied by another number to reach a desired number.

discriminator (see Figure 14). The discriminator's training is to judge whether a sample seems real or fake [56].

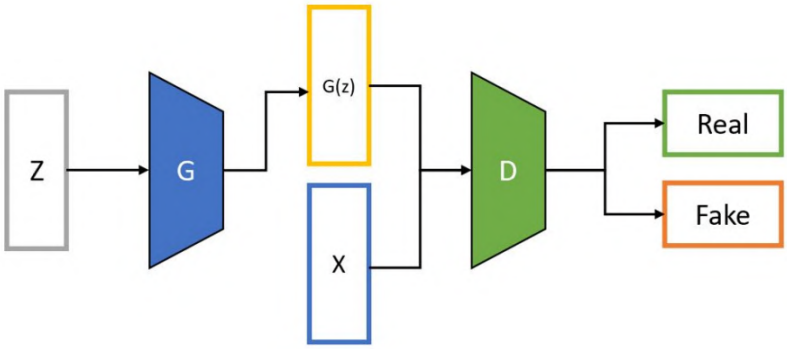

Figure 14: Generative adversarial networks architecture [57].

The generator network G_ϕ, with ϕ being the network parameters, and discriminator network D_θ, parameterised by θ, are simultaneously optimised using a gradient descent optimisation approach. The discriminator is trained in a supervised manner: the inputs consist of real and generated samples, and the output is the probability of an input being real or fake. This training process aims to minimise the cross-entropy cost function [56]:

$$J^{(D)} = -\mathbb{E}_{x\sim\mathbb{P}_r}\log D_\theta(x) - \mathbb{E}_{x\sim\mathbb{P}_g}\log(1 - D_\theta(x)). \tag{7}$$

The generator is a differentiable function that maps a sample from a noise prior, $\mathbf{z} \sim p(\mathbf{z})$, to the support of $\mathbb{P}_g$. The dimension of the noise vector $\mathbf{z}$ needs to be commensurate with the support of the convergence maps $\mathbb{P}_r$.

In the game theory formulation, the generator is trained to maximise eq. (7) known as the minimax game. The drawback of this formulation is that the gradients of the cost function vanish with respect to the generator parameters when the discriminator is winning. This makes difficult to train the generator using a gradient descent technique. Nevertheless, the first GAN work [56] was proposed flipping the target for the generator

$$J^{(G)} = -\mathbb{E}_{x\sim\mathbb{P}_g}\log D_\theta(x), \tag{8}$$

which corresponds to a cost function, known as the non-saturating game. This function is inspired in an heuristic that produces strong gradients during the generator training.

Since the creation of GAN, there have been a multitude of approaches for cost functions and other functional constrains such as conditioning.

Conditional GAN GANs may be extended to a conditional model, if both the generator and discriminator are designed to rely in an auxiliary information vector y, so called 'conditions', which may contain any kind of auxiliary information, such as class labels or data representing other modalities; Fig. 15 shows a typical conditional GAN. Conditioning may be performed by feeding y into both networks as an additional input layer, as well as using other input methods [58].

In the conditional generative adversarial networks (cGAN), which were first proposed in [59], the prior input noise $p_z(z)$, and y are combined in the generator through a joint hidden representation.

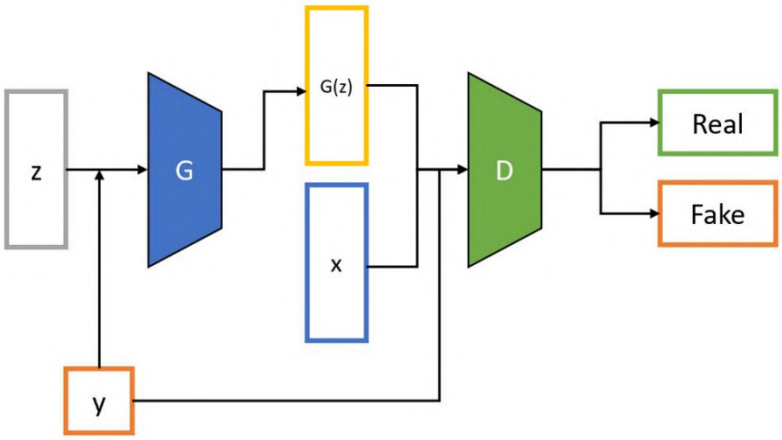

Figure 15: Conditional generative adversarial networks architecture [57].

The adversarial training framework allows for flexibility in its composition. In the discriminator, both x and y are presented as inputs to a discriminative function. The objective function of the two-player mini-max game is defined as [59]:

$$\min_{G}\max_{D} V(D,G) = \mathbb{E}_{x\sim p_{data}(x)}[\log D(x|y)] + E_{z\sim p_z(z)}[\log(1 - D(G(z|y)))]. \tag{9}$$

It is important to point out that subsequent studies, as in [58], [60], [61] have confirmed the feasibility of the generation of diverse, high-quality photo-realistic, and class-label consistent fake images using class-conditional GANs. Indeed, the works of [18] and [62] confirm a good performance with regression-label conditions.

3 Related Work

3.1 State of the Art on Conditional GANs

The first proposed cGANs [59], aimed to condition on some auxiliary information (class labels) the estimation of the probability distribution of images. Later studies [58], [60], [61] confirmed the viability of class-conditional GANs. However, an important drawback as described by Ding [63], is that these cGANs do not work well with continuous scalar conditions, known as regression labels, for image generation. This is mainly due to two problems:

- A cGAN have severe problems to estimate the image distribution on missing labels as it is usually trained to minimize the empirical versions of its losses; this is known as the empirical risk minimization (ERM) principle.
- In a cGAN, class labels are encoded by one-hot vectors or label embeddings, and then they are fed into both networks by hidden concatenation [59], an auxiliary classifier [58] or label projection [60]. Here the problem is that, as a precondition, the number of distinct labels is finite and known.

These problems make existing empirical losses and label input mechanisms of cGANs unsuitable for regression labels, as cGANs fail to generate images conditional on regression labels [63]. Binning the continuous conditions in disjoint intervals is a naive approach to solve these problems, training the cGAN in the usual class-conditional manner (intervals are considered independent classes). Still, this approach has four disadvantages [63]:

- cGANs collapse easily with this approach.
- Estimation only may be done on the membership in an interval and not on the target label.
- A large interval width carries to a high label inconsistency.
- It does not consider inter-class correlation.

The work of [64], proposed a model called continuous conditional generative adversarial network (CcGAN), which includes novel empirical losses, the hard vicinal discriminator loss (HVDL) and the soft vicinal discriminator loss (SVDL), and a new label input mechanisms. CcGAN is an important approach, even tough it still presents some issues when used for certain applications. In particular, some experiments with physical parameters as regression labels using CcGAN do not correlate as desired.

To overcome this issue for the inverse design of airfoils, [62] presented an alternative design (see Figure 16), which implements features such as the inclusion of an estimator, the computation of *Lambert log exponential transition score* (LLETS), the Determinantal Point Processes (DPPs) kernel, and a loss that follows the performance conditioning. All of these additions aimed to provide a better representation of data for an accurate correspondence between synthetic data with the expected real-world data.

This is an important progress, but still it is limited, as it only works for exclusively one parameter, and these kind of problems usually require more than one parameter to produce synthetic data, as mentioned by [62] and [18]. The approach of the solution proposed by [18] did not focus on

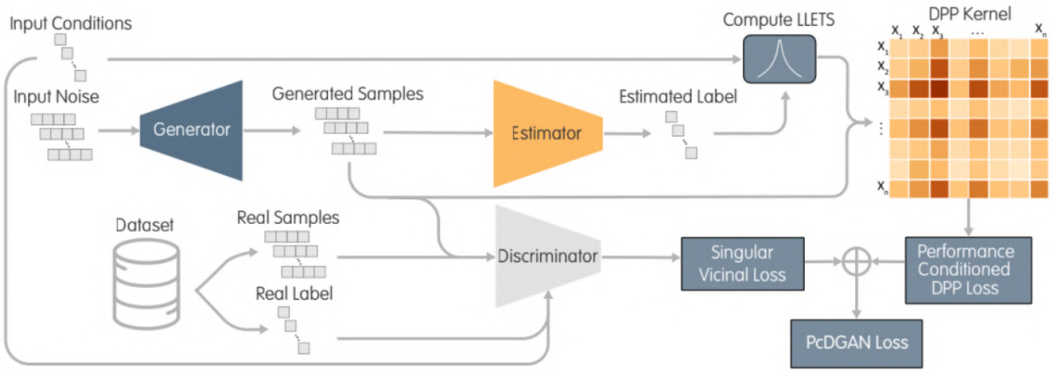

Figure 16: PcDGAN architecture [62].

changing the GAN architecture to improve the performance on the physical wold, but in the design of a new entrances method (see Figure 17), in order to give the model the ability of introduce more than one regression label as conditions. The authors obtained good results, but they missed the part of making an effort to improve an architecture to work for other problems of the physical world.

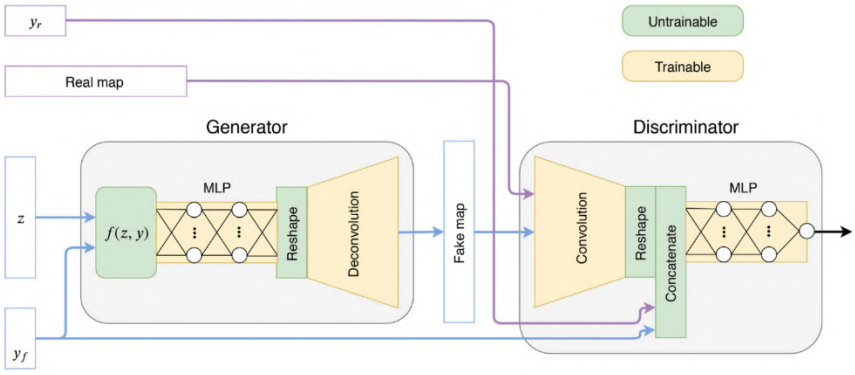

Figure 17: WGAN with new entrance method [18].

Table 1 shows some details of interest for this thesis work from those approaches.

3.1.1 Evaluation Metrics

The evaluation metrics being used in this research are considered from two origins: the similarity metrics traditionally used for computer vision and the summary statistics used in cosmology. Following the works of [18], [67], [68], the evaluation metrics considered are:

- The power spectral density C_l, which describes how strongly the maps are correlated as a function of pixel separation l [68].
- The distribution of the map pixels N_{pixels}, compared using Wasserstein-1 distance and histograms.

Table 1: A comparison of SOTA works in Conditional GANs

Work	**Model and main characteristics**	**Architecture Focused on Physics**	**Continuous Conditional Parameters**
GAN [65]	DCGAN + Interpolation afterwards	No	No
CGAN [66]	U-Net + CGAN	No	Yes (1 parameter) ?
CcGAN [64]	CGAN modified (Vicinal Loss + New Entrance Method)	No	Yes (1 parameter)
PcDGAN [62]	CcGAN modified (Estimator + New Entrance Method)	Yes	Yes (1 parameter)
Conditional WGAN [18]	CGAN modified (WGAN + New Entrance Method)	No	Yes (2 parameters)

- The distribution of the map peaks N_{peaks}, which describes the distribution of values at the local maxima of the map, also compared using Wasserstein-1 distance and histograms.
- The bispectrum B_l, which describes the three-point correlation of the folded triangles of different size.
- Minkowski functionals, which are morphological measures of the map, and they consist of three functions: V_0, which describes the area of the islands after thresholding of the map at some density level, V_1, their perimeter, and V_2, their Euler characteristic (their number count minus the number of holes) [69].
- The Pearson's correlation matrices Rll' between the C_l of maps of different cosmological models.
- The Multi-Scale Structural Similarity Index (MS-SSIM) [58], [70], which is an image similarity measure commonly used in computer vision.
- The Fréchet Distance between the output of a CNN regressor trained to predict Ω_m and σ_8 [71].

3.2 Generative Approaches on Cosmology

Cosmology has a great challenge while needing to generate simulations of theoretical models. In the era of precision cosmology, an important tool for studying the evolution of LSS are N-body simulations. Modelling the growth of structure at those scales involves non-linear physics that cannot be accurately described analytically. Modern cosmological simulations are highly realistic and extremely complex and may include galaxy evolution, feedback processes, massive neutrinos, weak lensing, and many other effects [65].

Currently the creation of each simulated virtual universe requires an extremely computationally expensive simulation on High Performance Computing (HPC) resources. That makes direct application of Markov chain MonteCarlo [72] or similar Bayesian methods prohibitively expensive to first estimate cosmological parameters required for these kind of simulations.

Lately, ML techniques have been explored in cosmology, including applications as an alternative to the traditional emulation methods. For instance, deep learning has been used to accurately predict non-linear structure formation [73]. Similarly, generative adversarial networks (GANs) [56]; and variational autoencoders (VAEs) [74]; have been used to produce novel realistic cosmic web 2D projections [55], weak lensing maps [3] and to perform dark energy model selection [75].

Weak gravitational lensing mass maps are considered to be one of the most powerful tools of extracting information from small scales. It probes both the geometry of the Universe and the growth of structure [76]. The shearing and magnification of background luminous sources by gravitational lensing allows us to reconstruct the matter distribution along the line of sight [55].

Only two levels of Generative adversarial networks (GAN) have recently been applied as a novel emulation technique for Large Scale Structure of the Universe (LSS) simulations. These models have the potential to make a dramatic change in the field of scientific simulations, but their performance for scientific applications needs to be studied. To do so, multiple contributions have leveraged the recent advances in the field of DL to aid the generation of cosmological simulations [3], [55], [68].

Like any algorithm, GAN has limitations, for example an unstable training procedure and the inherent randomness of the results produced. In [68], authors presented several techniques commonly used in computational learning to address these limitations on cosmological problems. They applied a latent space interpolation technique to control the results produced by the algorithm as an alternative to the use of Conditional generative adversarial networks (cGANs) [59].

Knowing that GANs have demonstrated their potential and that most existing GAN approaches produce simulations for a fixed value of cosmological parameters that limits their practical applicability, one paper proposed a new conditional GAN model, named Conditional Wasserstein GAN [18]. Their results showed that their conditional GAN may interpolate within the space of simulated cosmologies, and it is able to generate maps anywhere within this space with good visual quality and high statistical accuracy.

Table 2: Generative approaches used in cosmology.

Technique	**Tracer**
DNN - Deep Density Displacement Model (D3M) [73]	nonlinear LSS of the Universe with the Zel'dovich Approximation (ZA)
DCGAN [55]	weak lensing convergence maps
GAN, WGAN [3]	2D cosmic web
VAE+GAN [75]	dark energy reconstruction and model selection
GAN+interpolation [68]	weak lensing convergence maps and dark matter overdensity
WGAN, multiscale GAN, upscaling GAN conditioned [11]	2D and 3D cosmic web
CWGAN [18]	sky convergence maps

3.3 State of the Art Estimators on Vision Tasks

To date, convolutional neural networks (CNN) have been used for computer vision tasks. Although CNNs have managed to achieve far greater accuracy than other previous approaches, they present some shortcomings. CNNs were built for image classification, which is obtained by using successive layers of convolutions and pooling.

The pooling layer in a convolutional block is used to reduce the data dimension, as well as to achieve *spatial invariance* (regardless of where the object is placed in the image, it identifies the object and classifies it). Even though this is a powerful concept, it presents some drawbacks [53], e.g, while performing pooling, a CNN tends to lose a lot of information that is particularly useful, while performing tasks such as image segmentation and object detection. When the pooling layer loses the required spatial information about the rotation, location, scale, and different positional attributes of the object, the process of object detection and segmentation becomes very difficult.

While modern CNN architecture has managed to reconstruct the positional information using various advanced techniques, they are not completely accurate, and reconstruction itself is a tedious process [53]. Another drawback of the pooling layer is that if the position of the object is slightly changed, the activation does not seem to change with the same proportion, which leads to good accuracy in terms of image classification but poor in performance, if the user requires to locate exactly where the object is in an image.

To overcome this problem involving the rotational relationships in images, Capsnets emerged as a new model approach inspired in neuroscience (see section 2.2.4). CapsNets design incorporates dynamic routing algorithms to estimate features of objects like pose (position, size, deformation, velocity, etc). This approach proposes a new form of connection among neurons, and another architecture for the artificial neuron itself, called capsule[19], offering new interesting ways to design ANN architectures.

Another approach is the current SOTA GDL architectures. In recent years, variants of GNNs such as graph convolutional network (GCN), graph attention network (GAT), graph recurrent network (GRN) have demonstrated ground-breaking performances on many deep learning tasks [51]. In deep CNNs we rely on neighbouring pixels likely correlating to one another. To reproduce a similar setting on graphs, we need to take into account a reformulation of the "closeness". GDL approaches perform their tasks directly on the meshed structures or transform them into point clouds, achieving good performance and runtime [52]. The GDL constructive procedure provides a way to incorporate prior physical knowledge into neural architectures, additionally including a way to create new GNN model architectures [1].

3.4 Recovering BAO

Galaxy surveys encode cosmological information through effects like BAO. This pattern in the clustering of galaxies is complementary to other probes, such as SNIa and CMB. The most prominent use of standard rulers[12] is to measure the expansion rate of the Universe and its evolution using cosmic time, hence constraining the relative amounts between dark matter and dark energy.

The statistical uncertainties of BAO parameters are inherited from the uncertainty in the two-point statistics of galaxy clustering, which consist of two components: the shot noise and the cosmic variance [32]. While the former can be suppressed by increasing the density of tracers in the

[12] A standard ruler is a scale that is fixed through the evolution of the universe

observation, the latter cannot be reduced statistically, as long as only one tracer is observed, because of the limited number of pairs on large scales. However, if at least two tracers with an overlapping cosmic volume are observed, the cosmic variance can be reduced to some extent. This 'multi-tracer method' works if at least two differently biased tracers of the same underlying dark matter field are available. By comparing the galaxy clustering of these tracers, one is able to determine a ratio between the 'effective biases'. The validity of this method has been extensively studied [32]. Alternatively, one can combine tracers observed by different surveys. This is the approach chosen by SDSS or BOSS.

BOSS have measured the BAO scale at $z = 0.3$ and $z = 0.55$ using the galaxies distribution [77]. It has also measured the distribution of quasar absorption lines at $z = 2.5$, producing a measurement of the angular diameter distance, and a measure of the cosmic expansion rate $H(z)$.Nevertheless, BOSS is missing far infrared-view of dust-obscured star formation galaxies and AGN. In this regard, the TolTEC LSS survey is expected to provide the necessary detection and statistics to determine properties of the halos hosting dusty star forming galaxies, also known as SMGs, potentially providing information of the underlying galaxy distribution and statistical properties [9].

Furthermore, the imprint of non-linear processes on small angular scales cannot be captured with traditional analysis tools like the two-point statistics [78]. Multiple inference methods have been proposed to extract more details based on higher-order statistics, peak statistics, Minkowski functionals and recently CNNs [79].

For the reconstruction of the BAO signal and similar cosmology tasks, such as estimating cosmological parameters, there are recent approaches using ANNs and CNNs. For instance, the work of [80] estimated the values of cosmological parameters directly from the distribution of matter, using deep 3D CNNs for volumetric representation of dark-matter simulations. They also applied a distribution regression framework, showing that ML techniques are comparable to them and that they outperformed the results in some cases, as in maximum-likelihood point estimates using cosmological models. The work of [78], showed an improved CNN that obtained significantly better estimates of the Ω_m and σ_8 cosmological parameters from simulated weak lensing convergence maps, when compared to state-of-art methods. They also argued that their proposal is free of systematic bias. Such network exploits information in the gradients around peaks, that used to construct an easy-to-understand and robust peak-counting algorithm based on the steepness of peaks, instead of their heights. The proposed scheme is even more accurate than a ANN on high-resolution noiseless maps. With shape noise and lower resolution, its relative advantage deteriorates, but it remains more accurate than peak counting.

Other works, for example the one in [81], presented a method to reconstruct the initial conditions of the Universe using observed galaxy positions and luminosities under the assumption that the luminosities can be calibrated with weak lensing to give the mean halo mass, relying on following the gradients of forward model and since the standard way to identify halos is non-differentiable and results in a discrete sample of objects. Such work proposed a framework to model the halo position and mass field starting from the non-linear matter field using ANNs. It was shown that such reconstruction improves over the standard reconstruction. For BAO, the gains are relatively modest because BAO is dominated by large scales where standard reconstruction suffices, improving upon it by $\sim$ 15–20 percent in terms of error on BAO peak as estimated by Fisher analysis at $z = 0$.

A work specifically focused on recovering the BAO signal and reconstructing the initial condition based on the local density distribution in the configuration space, use a CNN not sensitive to cosmological parameters and works on cosmologies close to that of their training set [79]. It was

considered that this new scheme will possibly help to extract more information from the detected galaxy populations and those to be detected in the future[79]. These approaches will be used in this research as reference to determine the way to manage the data (See table 3).

Table 3: A comparison of SOTA works in BAOs recovering.

Technique	**Tracer used and year**
SDSS	I-II Nerby Galaxies (1998-2008)
BOSS	Old Red Galaxies (2008-2014)
eBOSS	Old Red Galaxies (2014-2019)
eBOSS	Young Blue Galaxies (2014-2019)
eBOSS +SDSS	I-II Quasars (1998-2019)
eBOSS + BOSS	Lyman-α (2008-2019)
CNN [81]	Simulation - Galaxy Light-Matter Connection (2018)
CNN [79]	Simulation - Density Matter Field (2020)

4 Research Proposal

4.1 Problem Statement

The problem tackled in this research is to create a novel DL framework to improve conditional data generation and pattern recognition, this to obtain a new approach suitable to perform an accurate detection of the characteristic separation and amplitude of the Baryon Acoustic Oscillations (BAO) from the submillimetre galaxy (SMG) population maps obtained by the LMT TolTEC camera at redshift $z = 2 - 3$.

The issues turn around the census about to be obtained by TolTEC of galaxy population at $z = 2 - 3$, which potentially could become a new tracer of matter distribution. This comes with the following problems: even though the data to be obtained is the most important survey of its kind, the detection of BAO would be difficult, as the number of galaxies surveyed may be insufficient, being necessary to complement the dataset with synthetic data that mimics the real data, and additionally to this, it is important to have new alternative approaches to obtain an improved measure of BAO, which could take advantage of the detection of a variety of intrinsic relationships on the images of the SMG population maps.

Considering a DL approach, the specific issues in our work are:

- The generation of synthetic data guided by conditions represented as regression labels. In this case, the goal is the generation of galactic distribution maps having as conditions the values of the matter density Ω_m, and the matter clustering strength σ_8. Both conditions strongly impact on the evolution of structures, for a given source galaxy redshift distribution n(z).

- The improvement of an estimator to recover from images a signal present in a pattern. In this particular case, the BAO signal over the submillimetre galaxy population maps.

4.2 Motivation and Justification

Cosmology has encountered a major challenge in data generation, processing and analysis. In the last few decades, astronomical observations of the early universe using CMB anisotropies, LSS distribution, luminosity distance inferred from supernova outbursts, and baryon acoustic oscillations (BAO) have provided ample evidence to support the cosmological model ΛCDM. Despite the success of this model, it is far from a complete description. Among other challenges, this model confronts that the physical properties of its components are mostly unknown. Indeed, recent measurements have found an inconsistency between the values obtained in the near and far universe. To solve these inconsistencies, it is necessary to perform the analysis of several observables over a wider redshift z interval, the careful control of the systematic effects associated with each measurement and the comparison of these observations with theoretical models. As we only have the observations of our own Universe, the way to experiment with evolution of universes following variations of cosmological models is through simulations.

The N-body techniques are very slow to run and computationally expensive, as they evolve the positions of particles over cosmic time in small time intervals and deep generative models have been proposed lately to simulate cosmological data, but most of the existing GAN approaches simulate for a fixed value of cosmological parameters (conditions), limiting a practical applicability of them

[18]. This last problem has been approached from various front lines, but it has shown not to have a trivial solution.

In pretty recent works, following the GAN approaches, class-conditional GANs has drawbacks, as described by the work of [63], which proposed Continuous Conditional GANs (CcGAN) attempting to solve these, but still having issues, as presented by [62], which points out its lack of good results when working with applications on physics and proposes an approach to face this problem. However, this work still misses the ability of handling more than one regression label as condition, which in turn is handled by the work of [18], but this last work misses the part of making an effort to improve an architecture to work for problems of the physical world. Hence, there are interesting possibilities to improve the generation of synthetic data for these purposes.

The statistical uncertainties of BAO parameters are inherited from the uncertainty in the two-point statistics of galaxy clustering. Furthermore, the imprint of non-linear processes on small angular scales cannot be captured with traditional analysis tools like the two-point statistics [78]. Multiple inference methods have been proposed to extract more details based on higher-order statistics, peak statistics, Minkowski functionals, and recently, CNNs [79].

Today, most of the DL approaches for this kind of problem have been focused on working on the improvement of CNNs. Even though CNNs have achieve greater accuracy, they still have shortcomings, as the pooling layer reduce the data dimension and achieve spatial invariance, but they tend to lose a lot of information particularly useful for image segmentation and object detection [53], such as spatial information about the rotation, location, scale, and different positional attributes of the object. This way, as described by [82], CNN treats the image as a grid or sequence structure, which is not flexible to capture irregular and complex objects. Instead, they proposed to represent the image as a graph structure and introduced a GNN architecture to extract graph-level feature for visual tasks. Another approach are the Capnets, which emerged as a new model approach incorporating dynamic routing algorithms to estimate features of objects as pose (position, size, deformation, velocity, texture, etc.) [19]. Here, capsules compute inputs and then ”encapsulate” the results into a small vector of highly informative outputs. A capsule could be considered a replacement for a common artificial neuron, which deals with scalars, while a capsule deals with vectors. These approaches represent an interesting manner to experiment and design new models that focuses in solving this kind of problems.

4.3 Research Questions

This research looks for answering the following questions:

- How to detect the characteristic separation and amplitude of the Baryon Acoustic Oscillations (BAO) from the submillimetre Galactic population maps by following a DL framework approach?
- How to improve a conditional generative adversarial network architecture, in order to use more than one continuous condition in the context of a truthfully data generation for galaxies representations by Baryon matter?
- How does the combination of properties of capsule networks models and geometric deep learning models affect in a DL model capability to improve features extraction and pattern recognition from images?

4.4 Hypothesis

A deep learning framework based on conditional deep generative models is able to produce synthetic images belonging to the original data probability distribution function, which complemented with a capsule network model will be able to recover the separation and amplitude of baryon acoustic oscillations (BAO), with a competitive performance compared to that of SOTA methods.

4.5 General Objective

To develop a new deep learning framework, able to recover the scale of baryon acoustic oscillations (BAO) at redshift z=2-3 using the sub-millimetre galaxy population as a tracer. This approach will be composed by a new conditional deep generative model, to produce enough synthetic images to complement the training data for a new capsule network model to extract image features with a competitive representation compared to that of other SOTA methods.

4.6 Specific Objectives

- To design a continuous conditional generative deep learning model to generate simulations of the distribution of galaxies in the optical and millimetre range of the electromagnetic spectrum, from a given set of cosmological parameters.
- To design a CapsNet architecture using these simulations and Baryon Oscillation Spectroscopic Survey (e-BOSS) observations as a base model, to obtain the BAO signal and characterise the systematics (errors) of this CapsNet results.
- To analyse the behaviour of the proposed framework with the legacy census observations of the large-scale structure obtained from TolTEC.

4.7 Expected Contributions

- A new continuous conditional generative deep learning model capable to receive more than one condition (a set of cosmological parameters), with a reliable distribution.
- A new DL architecture that combines properties of CapsNet and GDL allowing to extract a larger number of image features, compared to other extractors.
- A new cosmological tracer on the sub-millimetre region (A new application of the SMGs distribution as cosmological tracer).
- These new applications can contribute to mitigate the challenges to ΛCDM (solve tensions, discriminate between new cosmo models).

5 Methodology

Next we describe the methodology being followed in order to reach the objectives of this thesis (see Figure 18).

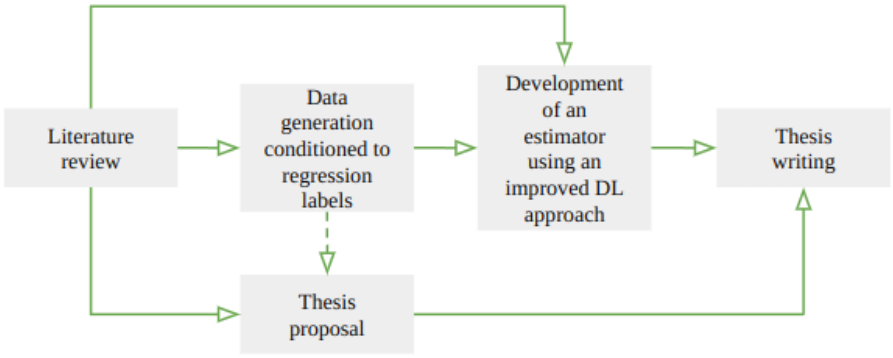

Figure 18: Process diagram of the thesis project.

5.1 Activities

5.1.1 Literature review

First, is required to have a clear view of the context in which the problems about to be solved are immersed. To do so, the following is being applied:

- Study the cosmology themes that are relevant to understand the patterns to be recognised and recovered.
- Study the SOTA on the traditional techniques used in cosmology to tackle these tasks.
- Study the deep learning approaches already developed for these purposes, analysing their strengths and weaknesses theoretically.
- Define the research proposal components (problem statement, motivation, justification, research questions, hypothesis, objectives and contributions), restricting these to the real expectations on the scope and limitations of this proposal.

5.1.2 Data generation conditioned to regression labels

From a given set of cosmological parameters, it is pretended to generate simulations of the distribution of galaxies in the optical and millimetre range of the electromagnetic spectrum using a GAN architecture. A preliminary idea of this model is showed in Fig. 19.

To solve this problem, the experiments to be developed are focused on reproducing different GAN models and testing a new GAN approach as follows:

- Reproduce the generation of convergence mass maps by a simple DCGAN.

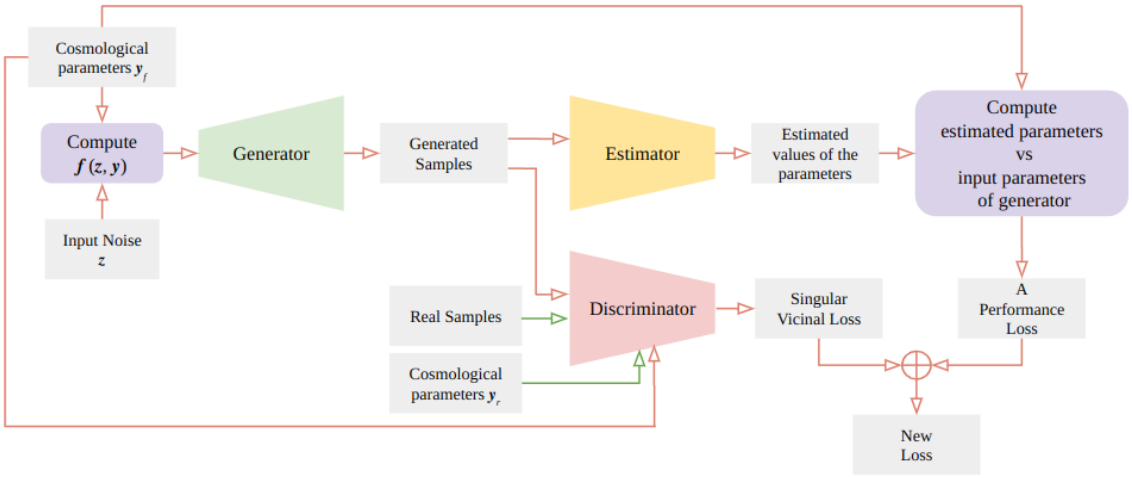

Figure 19: Preliminary architecture of the GAN model proposed.

- Reproduce the results of SOTA GANs conditioned on regression labels.
- Generate galaxies distribution maps conditioned on one regression label using SOTA models and analyse the advantages of each approach.
- Design a GAN model that outperforms SOTA models on using more than one regression label as condition. The idea is to design a better model, adapting the best features of the GANs for continuous conditioning. **Input:** *Cosmological model parameters (Ω_m and σ_8).* **Output:** *N-maps*.
- Analyse the performance of the new model and compare it with the existing SOTA approaches on a data set generated from a modified version of MultiDark-*Planck* (see Section 5.4), by using the evaluation metrics described on section 3.1.1.

5.1.3 Development of an estimator using an improved DL approach

Figure 20 shows a diagram of the complete system, which will be able to process the legacy census observations of the TolTEC large-scale structure through a new DL approach and find a value for the BAO scale.

The experiments to be developed are focused on testing different models and a new DL estimator model as follows:

- Implement and test an estimator model using a graph neural network (GNN) architecture. **Input:** *N-maps*. **Output:** *Scale and Amplitude of BAO*.
- Implement and test an estimator model using a capsule network (CapsNet) architecture. **Input:** *N-maps*. **Output:** *Scale and Amplitude of BAO*.
- Design an estimator model considering the best properties of a GNN architecture and a CapsNet architecture to be combined. It will be designed and trained to obtain the BAO signal and it will be characterise the systematics (errors), of the estimator results.

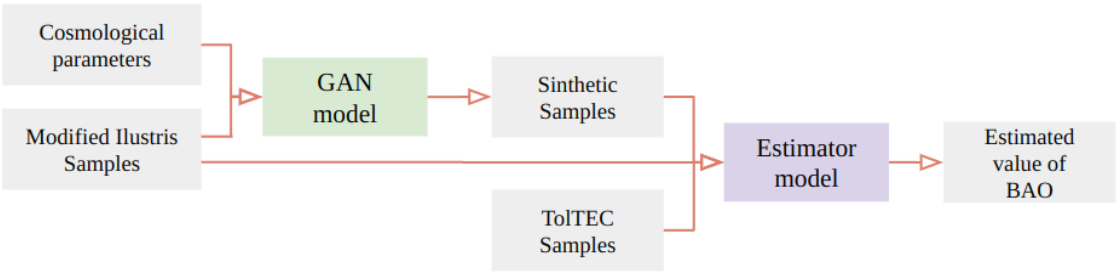

Figure 20: General diagram of the proposed system.

- Analyse the performance of the new model and compare it with the existing SOTA approaches (CNN's, GNN's and CapsNet's baselines).

5.2 Schedule

The activities to be carried out during this thesis work are shown in the Gantt chart of Figure 21.

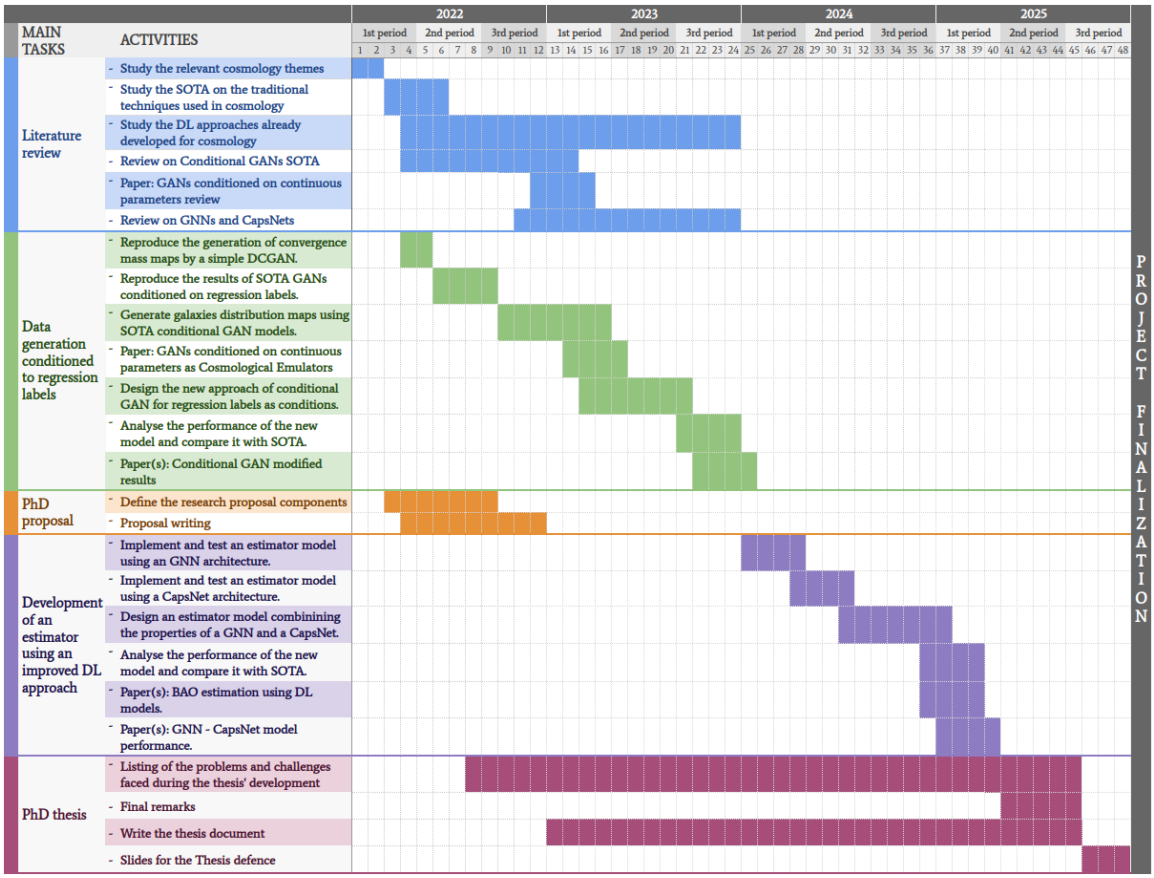

Figure 21: Activities schedule.

5.3 Development Tools

The main development environment considered for the experiments is Google Colaboratory PRO (Google Colab), which covers our processing needs by now. The following set-up is used:

- 27.3 gigabytes of RAM
- Tesla P100-PCIE GPU
- 167 GB of SSD hard drive space

For the implementation of experiments, it is considered to use Python 3, Keras 2.7 and TensorFlow 2. If necessary, PyTorch is another option that could be considered.

5.4 Data

The data for the GANs experiments consider two main datasets:

- The dataset provided by [83], without noise and intrinsic alignments. The simulation grid consists of 57 different cosmologies in the ΛCDM. This dataset contains labels corresponding to these 57 configurations with different values of both conditions (Ω_m and σ_8) that we expect to use in our own GAN model about to be designed. The projected matter distribution was pixelized into images of size 128 px $\times$ 128 px, which corresponds to $5^\circ \times 5^\circ$ of the sky. The resulting dataset consists of 57 sets of 12,000 sky convergence maps for a total of 684,000 samples.
- A dataset generated from the N-body simulation MultiDark-*Planck* [84], which is based on the ΛCDM model. This is a modified version populated with Sub-millimetric Galaxies, a project named "Simulations at 1.1/1.4/2.0 mm of the large-scale structure of the Universe to be studied with TolTEC", that was developed at INAOE [85].

On the other hand, the training and validation of the estimation experiments will use the data about to be obtained by TolTEC, described in Section 2.1.4, as well as the dataset of the modified MultiDark-*Planck* version and the synthetic data generated with our own GAN model.

5.5 Scope and Limitations

The aim is to design a GNN architecture capable of recovering the scale of BAO. The training and testing of this network will be performed by simulations and the extended Baryon Oscillation Spectroscopic Survey (e-BOSS) observations.

The upcoming TolTEC's observations of the legacy large-scale structure (LSS) will be analysed with the proposed framework and the possibility of detecting the BAO signal with the submillimetre galaxy (SMG) population will be assessed.

In case of a reliable detection of the BAO signal, the consistency of the BAO signal with CMB observations and BAO with other tracers will be analysed. Otherwise, other features and improvements for the design of future observations and strategies will be explored.

5.6 Publications Plan

It is expected to publish the following papers, a technical report of the thesis proposal and the thesis according to the activities and experiments developed during this project:

- GANs conditioned on continuous parameters: A review on theory and applications (2023).
- GANs conditioned on continuous parameters as Cosmological Emulators (2023). Proposed journal: astrophysical journal suplements.
- Article(s) on Conditional GAN modified (2024)
- Article(s) on GNN - CapsNet model performance (2025)
- Article(s) on BAO estimation using DL models (2025)
- Thesis (2025)

6 Preliminary Results

6.1 GAN experiments

The experiments developed so far followed the priority order presented before, focusing first on the deep generative models. Following a description of such experiments.

6.1.1 Reproducing CosmoGAN (A DCGAN as a Cosmological Emulator)

In order to understand the way that the GANs works in this context, the first experiment developed was a replication of CosmoGAN [55], to produce weak lensing convergence maps.

Dataset In order to train the GAN algorithm to produce realistic weak lensing convergence maps, we used a publicly available dataset. In particular, to test whether we can reproduce the original results from [55], we used the same dataset as in such work. This dataset consisted of 8000 weak lensing maps that were originally produced by running a GADGET2 [86] simulation with 512^3 particles in 240 Mpc h^{-1} box. To perform ray tracing, the GADGET weak lensing simulation pipeline was used. The simulation box was rotated multiple times for each ray tracing procedure, resulting in 1000 12 deg^2 maps per simulation box.

Model and training The architecture initially was adopted from the adaptation made by [55] of the Deep Convolutional GAN (DCGAN) [87], to meet the problem dimensions, and halved the number of filters.

In this architecture it is used a strided convolution and a transposed convolution, instead of pooling. The batch size was maintained in 64 maps. The minimisation loss process was implemented using Adam optimizers [74] with learning rate 0.00009 and β_1 = 0.5. The model of [55] uses learning rate 0.0002 and β_1 = 0.5, but this meant unstable training in our experiments.

We also modified the number of layers and we added one convolutional layer to each network. On the generator it was added on the penultimate layer and on the discriminator it was added next to the second layer. It was included a dropout of 0.2 on the discriminator, as dropout may improve generation quality, following the recommendation of [88].

The generator is a CNN, which in this case takes a random seed of size 64. The convolutional layers are transposed, with a kernel size of 5, strides of 2, and a random normal kernel initializer with standard deviation of 0.05. The batch normalization has a momentum of 0.9 and epsilon of 1e-5. The activation is ReLU in all layers, excepting the last one that is Tanh. Table 4 gives a summary of the generator architecture.

The discriminator is a DNN that receives as input a data vector and classifies it to indicate if the data is real or fake. The filters, kernel size, strides and batch normalization is the same as in the generator. The difference is that the generator uses convolutional layers, a truncated normal kernel initializer with standard deviation of 0.05 and an activation LeakyReLU with an α = 0.2. The last layer has a sigmoid activation. Table 5 shows a summary of the layers.

Evaluation Metrics The results obtained were analysed, in order to show that the GAN is not just mimicking a training dataset, but actually generating novel data. This work compares the generated maps samples with the validation dataset samples. The test consist of a 100 generated maps sample

Table 4: Generator's model.

Layer (type)	Activation	Output Shape	Param. number
InputLayer	-	64	0
Dense	-	32,768	2,129,920
Reshape	-	8, 8, 512	0
Conv2DTranspose	-	16, 16, 256	3,277,056
BatchNormalization	ReLU	16, 16, 256	1,024
Conv2DTranspose	-	32, 32, 128	819,328
BatchNormalization	ReLU	32, 32, 128	512
Conv2DTranspose	-	64, 64, 64	204,864
BatchNormalization	ReLU	64, 64, 64	256
Conv2DTranspose	-	128, 128, 32	51,232
BatchNormalization	ReLU	128, 128, 32	128
Conv2DTranspose	Tanh	256, 256, 1	801
Total params: 6,485,121 Trainable params: 6,484,161 Non-trainable params: 960			

Table 5: Discriminator model.

Layer (type)	Activation	Output Shape	Number of Param.
InputLayer	-	256, 256, 1	0
Conv2D	-	128, 128, 64	1,664
BatchNormalization	LeakyReLU	128, 128, 64	256
Conv2D	-	64, 64, 128	204,928
BatchNormalization	LeakyReLU	64, 64, 128	512
Conv2D	-	32, 32, 256	819,456
BatchNormalization	LeakyReLU	32, 32, 256	1,024
Conv2D	-	16, 16, 512	3,277,312
BatchNormalization	LeakyReLU	16, 16, 512	2,048
Flatten	-	131,072	0
Dropout	-	131,072	0
Dense	-	1	131,073
Total params: 4,438,273 Trainable params: 4,436,353 Non-trainable params: 1,920			

comparison against 100 validation dataset sample, this by calculating the difference in the pixel intensity histogram[13] using the seaborn library [89]. the power spectra[14] was also calculated using

[13]The pixel intensity is a first order statistic studied through an histogram showing the pixel intensities distribution of a dataset against another dataset.

[14]Power spectrum (PS) is the amplitudes of the Fourier components of a map. It is used to constrain cosmological models by comparing simulated maps to real data from sky surveys.

the code of [68], as well as the Minkowski functionals[15] on a 2D grid were calculated using the library proposed in [69].

Results It was found that the GAN algorithm was not just mimicking the training data, agreeing with the results found in the previous work of [55]. Examples of maps from the validation and GAN generated datasets (see Fig. 22).

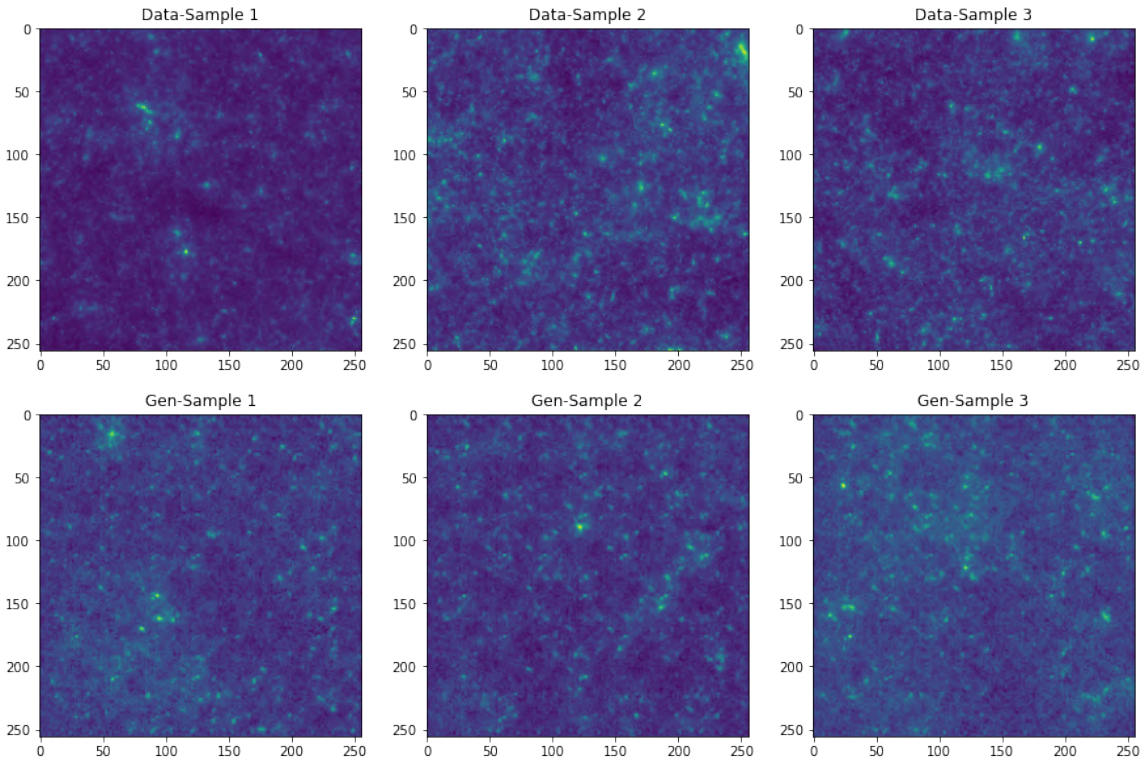

Figure 22: Images comparison: samples of validation data set vs generated data.

The histogram of pixel intensity distribution of an ensemble from the validation dataset was compared to that of generated maps of pixel intensities from the real simulation dataset. The generator succeeded on reproducing the convergence map structures, while evaluated by the power spectrum, confirming the statistical confidence reached by GANs to discover and reproduce such structures. Similarly, the findings with higher order statistics, in this case the Minkowski functionals, shown that GANs can be successful in reproducing the non-Gaussian structures in maps Fig. 23.

Discussion Another findings from this experiment showed that instability is a problem that requires to be solved using a fine tuning. Even a random seed for a new training can produce a different degree of performance on the results generated.

Another delicate issue is the number of epochs, as it can lead to mode collapse, that is a common problem on GANs; that was encountered multiple times in our training procedures. Training insta-

[15]The Minkowski functionals are used to study morphological features of fields that provide information of spatial correlation, object shapes and topology, allowing to capture non-Gaussian information on small scales.

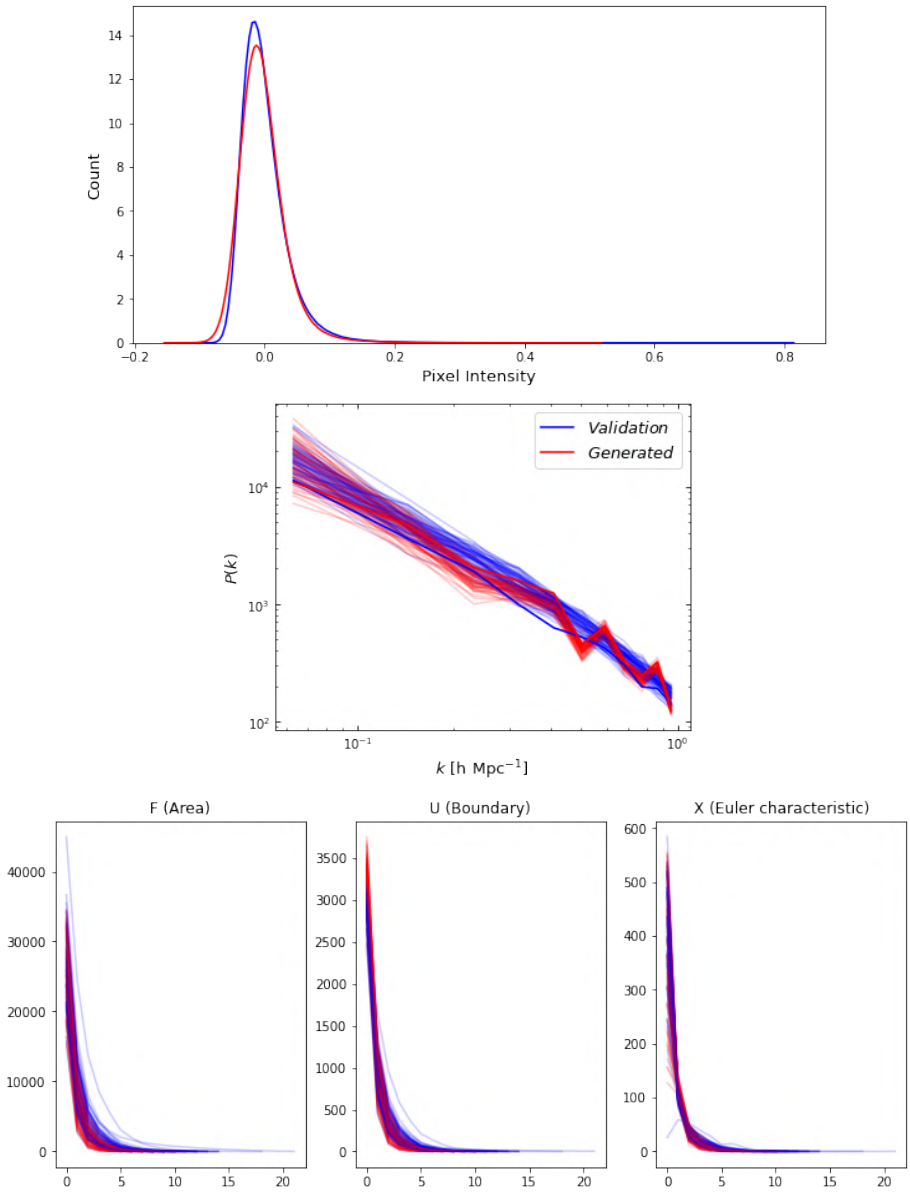

Figure 23: Blue line represents the validation data and red line the generated data. Top-left: Pixel intensity histogram for an ensemble of 100 weak lensing convergence maps. Top-right: Power Spectrum. Bottom: Minkowski Functionals of the comparison.

bility can be a challenging problem on any kind of GANs architecture, as reported by [55], [68]. Many epochs produced degeneration, or collapse on GANs, so adjustment of this was necessary, while developing this experiment. Adding layers to reduce forced dimensionality and it also helped to generate images with better fidelity. Using dropout on the discriminator and adjusting learning rate may stabilize GANs.

The metrics used in this work were used with a basic configuration, so it is important to tune these statistics, in order to get a better perception of the performance of GANs. Additionally it would be

useful to integrated some other statistics. For example, it may be used the extensive set of statistics for quantitative comparison of the N-body and GAN generated maps presented in [18].

6.1.2 Reproducing GANs Conditioned on Regression Labels

The next experiments developed were focused on testing the SOTA architectures that are considered to guide in the solution of the first task of this thesis work, specifically the Continuous Conditional GAN (CcGAN) [64] and the Performance Conditioned Diverse GAN (PcDGAN) [62].

Dataset To demonstrate the PcDGAN performance, the work of [62] created two simulated examples. There are six peaks in the conditioning space making the problem a multi-modal problem. For every input condition, there are points in all six modes and a method which can generate samples in all the modes is preferred.

For the first example, data is uniformly distributed within -0.6 and 0.6 of the origin in both directions, simulating a case where data is evenly covering the label space without any bias. For the second example a similar distribution exists for 50 percent of the data, while the other 50 percent of the data is located in a circle around one of the peaks. The second example is used to demonstrate the case where the data is distributed unevenly in the design space, which may lead to mode collapse. In both examples, 10,000 data points are used for training.

Models and training The testings of [64] and [62] were developed with the source code provided by [62], specifically designed to compare both architectures.

Evaluation Metrics The metrics proposed in [62] focused on measuring the performance over the continuous conditions labelling (likelihood score, label error and diversity). These metrics were used the same way to replicate the results expected.

Results The execution showed the same results (see Fig. 24) as the published paper [62].

Discussion The experiments with conditional GANs for continuous conditions where interesting, as they have shown the same performance reported on the synthetic datasets, and we could execute and experiment with these implementations without any problem. Just changes were made to adjust the source code to the Google Colab environment; it was not required any adaptation of the architecture presented for the source code provided by [62].

6.1.3 Generating galaxy distribution maps conditioned on one regression label

Dataset The dataset considered for these experiments is the one provided by [83] (see section 5.4), and used by [18] to train their model. This dataset contains labels corresponding to both conditions (Ω_m and σ_8) that we expect to use in our own GAN model about to be design.

Models and training Developing this experiment in particular is the main step to understand how a conditional GAN model may be improved to work with continuous conditions, and to find out the drawbacks of each one of the SOTA models [18], [62], [64] (described in section 3.1), in this regard.

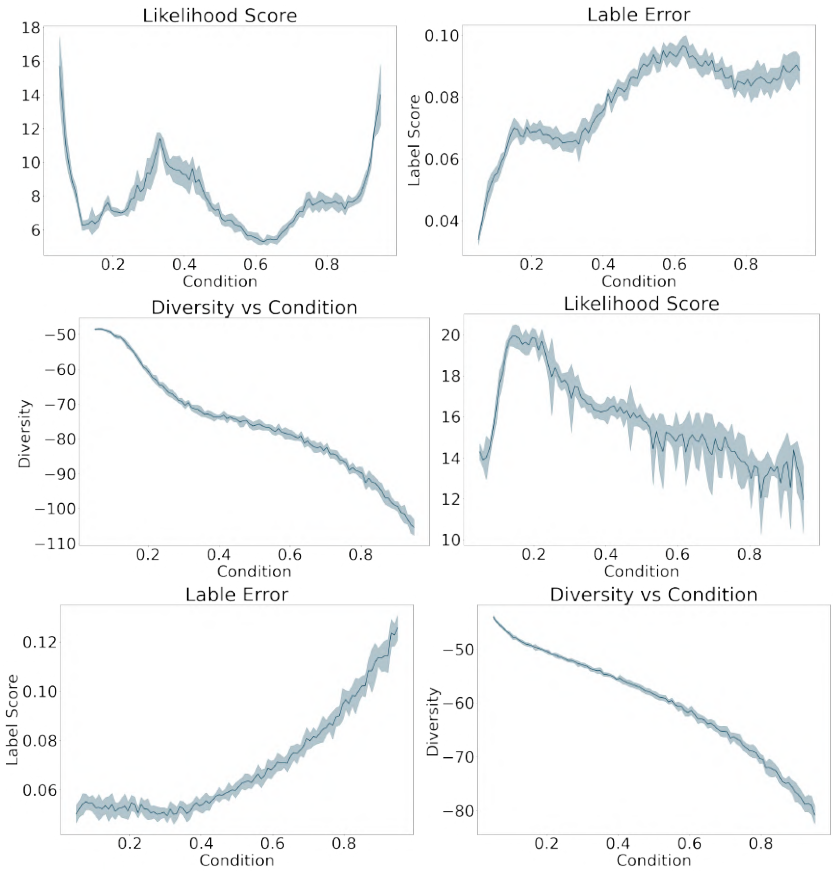

Figure 24: The plots indicate the mean and standard deviation of CcGAN (top) and PcDGAN (bottom) for all three metrics.

Evaluation Metrics The metrics proposed for these experiments are the ones proposed by [62] and [18], as described in Section 3.1.1, focused on measuring the performance over the continuous conditions labelling.

Results This part of the experiments is currently being carried out.

Discussion Unfortunately, we could not yet finish a successful execution using the previously named dataset. It seems that it will be required to develop more tuning, in order to adjust the hyperparameters properly. An issue that slowed down the development of these experiments was the implementation of a new source code, as the packages used by [18], was not compatible with the Google Colab platform.

7 Final Remarks

The use of GANs is considerable promising for cosmology applications, as they offers an alternative for fast generation of reliable synthetic data alternative. However, even though GANs can get images generated with a very high-fidelity, replicating pixel intensity, power spectrum and higher-order statistics, this is not a straightforward action, due to both the complexity in the statistics ruling the training data in this area, and its dependence on input continuous labels.

It is important to point out that in the last year it has been developed almost all the works on GANs dealing with continuous conditions, showing that this area is a trending nowadays [18], [62], [64], [66].

Indeed, the models for the estimation of the BAO values require to be improved. For this particular problem, and considering other new approaches suggested in the last five years for these kind of tasks, the new researches are exploring strategies different to that of the CNNs or typical ANNs, this to avoid the known limitations found in those models. This way, there are emerging new models such as CapsNets, or approaches to have new constructive procedures in the case of GDL.

Therefore, the main contribution of this thesis will be two new models: A conditional GAN model able to produce accurate synthetic data and a CapsNet model able to calculate BAO values, both with a competitive performance compared to the current SOTA applications.

Summarising, the work in this year has been focused on the documentary research developed to write the theoretical frame, the SOTA and the PhD proposal sections. Additionally, it has been developed some relevant experiments focused on clarifying the approach that will be taken for the first part of the framework proposed, themed on generative adversarial networks. These activities have conducted to two papers which are in process of being writing.

References

[1] M. M. Bronstein, J. Bruna, T. Cohen, and P. Veličković, "Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges," *arXiv:2104.13478 [cs, stat]*, May 2021, arXiv: 2104.13478. [Online]. Available: http://arxiv.org/abs/2104.13478 (visited on 03/11/2022).

[2] B. Ryden, *Introduction to Cosmology*. Cambridge University Press, 2017, ISBN: 9781107154834. [Online]. Available: https://books.google.com.mx/books?id=07WSDQAAQBAJ.

[3] A. C. Rodríguez, T. Kacprzak, A. Lucchi, *et al.*, "Fast cosmic web simulations with generative adversarial networks," en, *Computational Astrophysics and Cosmology*, vol. 5, no. 1, p. 4, Nov. 2018, ISSN: 2197-7909. DOI: 10.1186/s40668-018-0026-4. [Online]. Available: https://doi.org/10.1186/s40668-018-0026-4 (visited on 12/02/2021).

[4] E. Di Valentino, A. Melchiorri, and J. Silk, "Planck evidence for a closed Universe and a possible crisis for cosmology," en, *Nature Astronomy*, vol. 4, no. 2, pp. 196–203, Feb. 2020, Number: 2 Publisher: Nature Publishing Group, ISSN: 2397-3366. DOI: 10.1038/s41550-019-0906-9. [Online]. Available: https://www.nature.com/articles/s41550-019-0906-9 (visited on 03/11/2022).

[5] A. Loeb, "The reionization of cosmic hydrogen by the first galaxies," in *Adventures In Cosmology*, World Scientific, 2011, pp. 41–88.

[6] A.-M. Weijmans, M. Blanton, A. S. Bolton, J. Brownstein, M. J. Raddick, and A. Thakar, "The Challenges of a Public Data Release: behind the Scenes of SDSS DR13," in *Astronomical Data Analysis Software and Systems XXVI*, M. Molinaro, K. Shortridge, and F. Pasian, Eds., ser. Astronomical Society of the Pacific Conference Series, vol. 521, Oct. 2019, p. 177.

[7] G.-B. Zhao, Y. Wang, A. J. Ross, *et al.*, "The extended Baryon Oscillation Spectroscopic Survey: a cosmological forecast," *Monthly Notices of the Royal Astronomical Society*, vol. 457, no. 3, pp. 2377–2390, Feb. 2016, ISSN: 0035-8711. DOI: 10.1093/mnras/stw135. eprint: https://academic.oup.com/mnras/article-pdf/457/3/2377/8000992/stw135.pdf. [Online]. Available: https://doi.org/10.1093/mnras/stw135.

[8] A. Becker. "Dark energy spectroscopic instrument (DESI) creates largest 3d map of the cosmos -," News Center. (Jan. 13, 2022), [Online]. Available: https://newscenter.lbl.gov/2022/01/13/dark-energy-spectroscopic-instrument-desi-creates-largest-3d-map-of-the-cosmos/ (visited on 12/14/2022).

[9] LMT, *The TolTEC Camera*. [Online]. Available: http://toltec.astro.umass.edu/science_lss.php (visited on 04/22/2022).

[10] *LMT - Large Millimeter Telescope Alfonso Serrano*, en. [Online]. Available: http://lmtgtm.org/?lang=es (visited on 07/05/2022).

[11] N. Perraudin, A. Srivastava, A. Lucchi, T. Kacprzak, T. Hofmann, and A. Réfrégier, "Cosmological n-body simulations: A challenge for scalable generative models," *Computational Astrophysics and Cosmology*, vol. 6, no. 1, p. 5, Dec. 19, 2019, ISSN: 2197-7909. DOI: 10.1186/s40668-019-0032-1. [Online]. Available: https://doi.org/10.1186/s40668-019-0032-1 (visited on 12/14/2022).

[12] F. Villaescusa-Navarro, *Features — Quijote simulations 0.1 documentation*, 2020. [Online]. Available: https://quijote-simulations.readthedocs.io/en/latest/features.html (visited on 07/05/2022).

[13] C. Howlett, M. Manera, and W. J. Percival, "L-PICOLA: A parallel code for fast dark matter simulation," *Astronomy and Computing*, vol. 12, pp. 109–126, 2015, ISSN: 2213-1337. DOI: https://doi.org/10.1016/j.ascom.2015.07.003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2213133715000700.

[14] V. Springel, R. Pakmor, O. Zier, and M. Reinecke, "Simulating cosmic structure formation with the gadget-4 code," *Monthly Notices of the Royal Astronomical Society*, vol. 506, no. 2, pp. 2871–2949, Jul. 2021, ISSN: 0035-8711. DOI: 10.1093/mnras/stab1855. eprint: https://academic.oup.com/mnras/article-pdf/506/2/2871/39271725/stab1855.pdf. [Online]. Available: https://doi.org/10.1093/mnras/stab1855.

[15] J. R. Primack. "The cosmological supercomputer," IEEE Spectrum. Section: Aerospace. (Oct. 1, 2012), [Online]. Available: https://spectrum.ieee.org/the-cosmological-supercomputer (visited on 12/14/2022).

[16] S. Cheng, H.-R. Yu, D. Inman, Q. Liao, Q. Wu, and J. Lin, "Cube – towards an optimal scaling of cosmological n-body simulations," in *2020 20th IEEE/ACM International Symposium on Cluster, Cloud and Internet Computing (CCGRID)*, 2020, pp. 685–690. DOI: 10.1109/CCGrid49817.2020.00-22.

[17] M. Vogelsberger, S. Genel, V. Springel, *et al.*, "Introducing the Illustris Project: simulating the coevolution of dark and visible matter in the Universe,", vol. 444, no. 2, pp. 1518–1547, Oct. 2014. DOI: 10.1093/mnras/stu1536. arXiv: 1405.2921 [astro-ph.CO].

[18] N. Perraudin, S. Marcon, A. Lucchi, and T. Kacprzak, "Emulation of Cosmological Mass Maps with Conditional Generative Adversarial Networks," *Frontiers in Artificial Intelligence*, vol. 4, p. 66, 2021, ISSN: 2624-8212. DOI: 10.3389/frai.2021.673062. [Online]. Available: https://www.frontiersin.org/article/10.3389/frai.2021.673062 (visited on 11/18/2021).

[19] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, ser. NIPS'17, Long Beach, California, USA: Curran Associates Inc., 2017, pp. 3859–3869, ISBN: 9781510860964.

[20] M. M. Bronstein, J. Bruna, T. Cohen, and P. Veličković, *Geometric deep learning: Grids, groups, graphs, geodesics, and gauges*, 2021. DOI: 10.48550/ARXIV.2104.13478. [Online]. Available: https://arxiv.org/abs/2104.13478.

[21] Western Washington University. "Hubble's law — cosmic redshift — western washington university." (), [Online]. Available: https://www.wwu.edu/astro101/a101_hubble_redshift.shtml (visited on 12/24/2022).

[22] NASA. "WMAP- content of the universe." (), [Online]. Available: https://wmap.gsfc.nasa.gov/universe/uni_matter.html (visited on 01/16/2023).

[23] P. Schneider, *Extragalactic Astronomy and Cosmology: An Introduction*. Springer Berlin Heidelberg, 2014, ISBN: 978-3-642-54083-7. [Online]. Available: https://books.google.com.mx/books?id=kcrHBAAAQBAJ.

[24] W. L. Freedman, B. F. Madore, D. Hatt, *et al.*, "The carnegie-chicago hubble program. viii. an independent determination of the hubble constant based on the tip of the red giant branch*," *The Astrophysical Journal*, vol. 882, no. 1, p. 34, Aug. 2019. DOI: 10.3847/1538-4357/ab2f73. [Online]. Available: https://dx.doi.org/10.3847/1538-4357/ab2f73.

[25] L. Kazantzidis and L. Perivolaropoulos, "σ_8 Tension. is gravity getting weaker at low z? observational evidence and theoretical implications," in *Modified Gravity and Cosmology: An Update by the CANTATA Network*, E. N. Saridakis, R. Lazkoz, V. Salzano, *et al.*, Eds. Cham: Springer International Publishing, 2021, pp. 507–537, ISBN: 978-3-030-83715-0. DOI: 10.1007/978-3-030-83715-0_33. [Online]. Available: https://doi.org/10.1007/978-3-030-83715-0_33.

[26] A. Pradhan, "Thermal history of the universe," p. 44, 2018.

[27] Swinburne University of Technology. "Baryonic matter — COSMOS." (), [Online]. Available: https://astronomy.swin.edu.au/cosmos/B/Baryonic+Matter (visited on 12/14/2022).

[28] M. Roos, *Introduction to Cosmology*. Wiley, 2015, ISBN: 9781118923290. [Online]. Available: https://books.google.com.mx/books?id=RkgZBwAAQBAJ.

[29] ESA. "Planck and the cosmic microwave background." (), [Online]. Available: https://www.esa.int/Science_Exploration/Space_Science/Planck/Planck_and_the_cosmic_microwave_background (visited on 01/15/2023).

[30] Monica Young. "Dark energy BOSSes around the universe," Sky & Telescope. (Apr. 12, 2012), [Online]. Available: https://skyandtelescope.org/astronomy-news/dark-energy-bosses-around-theuniverse/ (visited on 01/15/2023).

[31] N. GMS, *GMS: Baryon Acoustic Oscillations*, en, Nov. 2020. [Online]. Available: https://svs.gsfc.nasa.gov/13768 (visited on 04/22/2022).

[32] Y. Wang and G.-B. Zhao, "A brief review on cosmological analysis of galaxy surveys with multiple tracers," *Research in Astronomy and Astrophysics*, vol. 20, no. 10, p. 158, Oct. 2020. DOI: 10.1088/1674-4527/20/10/158. [Online]. Available: https://doi.org/10.1088/1674-4527/20/10/158.

[33] W. L. Freedman and B. F. Madore, "The hubble constant," *Annual Review of Astronomy and Astrophysics*, vol. 48, no. 1, pp. 673–710, 2010. DOI: 10.1146/annurev-astro-082708-101829. eprint: https://doi.org/10.1146/annurev-astro-082708-101829. [Online]. Available: https://doi.org/10.1146/annurev-astro-082708-101829.

[34] W. Hu, N. Sugiyama, and J. Silk, "The physics of microwave background anisotropies," *Nature*, vol. 386, no. 6620, pp. 37–43, Mar. 1997, Number: 6620 Publisher: Nature Publishing Group, ISSN: 1476-4687. DOI: 10.1038/386037a0. [Online]. Available: https://www.nature.com/articles/386037a0 (visited on 12/26/2022).

[35] SDSS. "No need to mind the gap: Astrophysicists fill in 11 billion years of our universe's expansion history — SDSS — press releases." (Jul. 19, 2020), [Online]. Available: https://www.sdss.org/press-releases/no-need-to-mind-the-gap/ (visited on 04/22/2022).

[36] S. Dodelson and F. Schmidt, "11 - probes of structure: Tracers," in *Modern Cosmology (Second Edition)*, S. Dodelson and F. Schmidt, Eds., Second Edition, Academic Press, 2021, pp. 295–323, ISBN: 978-0-12-815948-4. DOI: https://doi.org/10.1016/B978-0-12-815948-4.00017-6. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780128159484000176.

[37] T. P. Downes, "The population of submillimeter galaxies and its impact on the detection of the Sunyaev-Zel'Dovich Effect," Ph.D. dissertation, University of Chicago, Jan. 2009.

[38] ESA. "ESA science & technology - hunting galaxy clusters with planck – the sunyaev-zel'dovich effect." (), [Online]. Available: https://sci.esa.int/web/planck/-/48227-hunting-galaxy-clusters-with-planck-the-sz-effect (visited on 01/16/2023).

[39] M. McCrackan, Z. Ma, N. S. DeNigris, *et al.*, "The TolTEC camera: the citlali data reduction pipeline engine," in *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, ser. Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, vol. 12189, Aug. 2022, 121891H, 121891H. DOI: 10.1117/12.2629095.

[40] R. Rew, G. Davis, S. Emmerson, *et al.*, *Unidata NetCDF*, in collab. with W. Fisher, R. Rew, W. Fisher, U. C. F. A. R. (U. C. P. (Unidata, U. C. F. A. R. (U. C. P. (Unidata, and E. S. R. L. National Oceanic {and} Atmospheric Administration, Language: en Medium: application/java-archive,application/gzip,application/tar, 1989. DOI: 10.5065/D6H70CW6. [Online]. Available: http://www.unidata.ucar.edu/software/netcdf/ (visited on 01/17/2023).

[41] NASA. "FITS primer." (), [Online]. Available: https://fits.gsfc.nasa.gov/fits_primer.html (visited on 01/17/2023).

[42] T. M. Mitchell, *Machine Learning* (McGraw-Hill series in computer science). New York: McGraw-Hill, 1997, 414 pp., ISBN: 978-0-07-042807-2.

[43] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[44] IBM. "What is supervised learning? — IBM." (), [Online]. Available: https://www.ibm.com/topics/supervised-learning (visited on 01/02/2023).

[45] I. Goodfellow, J. Pouget-Abadie, M. Mirza, *et al.*, "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, Oct. 2020, ISSN: 0001-0782. DOI: 10.1145/3422622. [Online]. Available: https://doi.org/10.1145/3422622.

[46] Y. Bengio and Y. LeCun, "Scaling learning algorithms towards ai," 2007.

[47] O. Delalleau and Y. Bengio, "Shallow vs. deep sum-product networks," in *NIPS*, 2011.

[48] R. Pascanu, C. Gulcehre, K. Cho, and Y. Bengio, "How to construct deep recurrent neural networks: Proceedings of the second international conference on learning representations (iclr 2014)," in *International Conference on Learning Representations*, 2014.

[49] G. Montúfar, R. Pascanu, K. Cho, and Y. Bengio, "On the Number of Linear Regions of Deep Neural Networks," *arXiv e-prints*, arXiv:1402.1869, arXiv:1402.1869, Feb. 2014. arXiv: 1402.1869 [stat.ML].

[50] M. Bronstein. "Geometric foundations of deep learning," Medium. (Aug. 10, 2022), [Online]. Available: https://towardsdatascience.com/geometric-foundations-of-deep-learning-94cdd45b451d (visited on 06/01/2023).

[51] J. Zhou, G. Cui, S. Hu, *et al.*, "Graph neural networks: A review of methods and applications," *AI Open*, vol. 1, pp. 57–81, 2020, ISSN: 2666-6510. DOI: https://doi.org/10.1016/j.aiopen.2021.01.001. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666651021000012.

[52] A. Pearce, A. Wiltschko, B. Sanchez-Lengeling, and E. Reif, "A gentle introduction to graph neural networks," *Distill*, vol. 2021, N/A, 2021. [Online]. Available: https://distill.pub/2021/gnn-intro/.

[53] V. Kurama, *Introduction to Capsule Networks*, en, Jul. 2020. [Online]. Available: https://blog.paperspace.com/capsule-networks/ (visited on 03/12/2022).

[54] H. Liao, *CapsNet-tensorflow*, original-date: 2017-10-28T04:20:36Z, Jan. 16, 2023. [Online]. Available: https://github.com/naturomics/CapsNet-Tensorflow (visited on 01/17/2023).

[55] M. Mustafa, D. Bard, W. Bhimji, Z. Lukić, R. Al-Rfou, and J. M. Kratochvil, "CosmoGAN: Creating high-fidelity weak lensing convergence maps using Generative Adversarial Networks," en, *Computational Astrophysics and Cosmology*, vol. 6, no. 1, p. 1, May 2019, ISSN: 2197-7909. DOI: 10.1186/s40668-019-0029-9. [Online]. Available: https://doi.org/10.1186/s40668-019-0029-9 (visited on 11/18/2021).

[56] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, *et al.*, "Generative Adversarial Networks," *arXiv:1406.2661 [cs, stat]*, Jun. 2014, arXiv: 1406.2661. [Online]. Available: http://arxiv.org/abs/1406.2661 (visited on 12/02/2021).

[57] M. Ahang, M. Jalayer, A. Shojaeinasab, O. Ogunfowora, T. Charter, and H. Najjaran, "Synthesizing rolling bearing fault samples in new conditions: A framework based on a modified cgan," *Sensors*, vol. 22, no. 14, 2022, ISSN: 1424-8220. DOI: 10.3390/s22145413. [Online]. Available: https://www.mdpi.com/1424-8220/22/14/5413.

[58] A. Odena, C. Olah, and J. Shlens, "Conditional Image Synthesis with Auxiliary Classifier GANs," en, in *Proceedings of the 34th International Conference on Machine Learning*, ISSN: 2640-3498, PMLR, Jul. 2017, pp. 2642–2651. [Online]. Available: https://proceedings.mlr.press/v70/odena17a.html (visited on 12/02/2021).

[59] M. Mirza and S. Osindero, "Conditional Generative Adversarial Nets," *arXiv:1411.1784 [cs, stat]*, Nov. 2014, arXiv: 1411.1784. [Online]. Available: http://arxiv.org/abs/1411.1784 (visited on 11/30/2021).

[60] T. Miyato and M. Koyama, "cGANs with Projection Discriminator," en, Feb. 2018. [Online]. Available: https://openreview.net/forum?id=ByS1VpgRZ (visited on 12/02/2021).

[61] A. Brock, J. Donahue, and K. Simonyan, "Large Scale GAN Training for High Fidelity Natural Image Synthesis," en, Sep. 2018. [Online]. Available: https://openreview.net/forum?id=B1xsqj09Fm (visited on 12/02/2021).

[62] A. Heyrani Nobari, W. Chen, and F. Ahmed, "PcDGAN: A Continuous Conditional Diverse Generative Adversarial Network For Inverse Design," en, in *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, Virtual Event Singapore: ACM, Aug. 2021, pp. 606–616, ISBN: 978-1-4503-8332-5. DOI: 10.1145/3447548.3467414. [Online]. Available: https://dl.acm.org/doi/10.1145/3447548.3467414 (visited on 11/28/2021).

[63] X. Ding, Y. Wang, Z. Xu, W. J. Welch, and Z. J. Wang, "CcGAN: Continuous Conditional Generative Adversarial Networks for Image Generation," en, Sep. 2020. [Online]. Available: https://openreview.net/forum?id=PrzjugOsDeE (visited on 11/30/2021).

[64] X. Ding, *Continuous Conditional Generative Adversarial Networks*, original-date: 2020-02-25T00:06:57Z, Nov. 2021. [Online]. Available: https://github.com/UBCDingXin/improved_CcGAN (visited on 11/30/2021).

[65] A. Tamosiunas, *A GAN algorithm to produce weak lensing convergence maps and cosmic web slices:* original-date: 2020-02-02T03:44:00Z, Jul. 2021. [Online]. Available: https://github.com/AndriusT/cw_wl_GAN (visited on 11/30/2021).

[66] D. Schaurecker, Y. Li, J. Tinker, S. Ho, and A. Refregier, "Super-resolving Dark Matter Halos using Generative Deep Learning," *arXiv:2111.06393 [astro-ph]*, Nov. 2021, arXiv: 2111.06393. [Online]. Available: http://arxiv.org/abs/2111.06393 (visited on 11/30/2021).

[67] M. Mustafa, *MustafaMustafa/cosmoGAN*, original-date: 2017-06-19T16:45:40Z, Jan. 2021. [Online]. Available: https://github.com/MustafaMustafa/cosmoGAN (visited on 11/30/2021).

[68] A. Tamosiunas, H. A. Winther, K. Koyama, D. J. Bacon, R. C. Nichol, and B. Mawdsley, "Investigating cosmological GAN emulators using latent space interpolation," *Monthly Notices of the Royal Astronomical Society*, vol. 506, no. 2, pp. 3049–3067, Sep. 2021, ISSN: 0035-8711. DOI: 10.1093/mnras/stab1879. [Online]. Available: https://doi.org/10.1093/mnras/stab1879 (visited on 11/18/2021).

[69] A. M. P. Boelens and H. A. Tchelepi, *Quantimpy*, original-date: 2021-09-14T08:41:47Z, Sep. 2021. [Online]. Available: https://github.com/ElsevierSoftwareX/SOFTX-D-21-00096 (visited on 11/30/2021).

[70] Z. Wang, E. Simoncelli, and A. Bovik, "Multiscale structural similarity for image quality assessment," in *The Thrity-Seventh Asilomar Conference on Signals, Systems Computers, 2003*, vol. 2, 2003, 1398–1402 Vol.2. DOI: 10.1109/ACSSC.2003.1292216.

[71] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, ser. NIPS'17, Long Beach, California, USA: Curran Associates Inc., 2017, pp. 6629–6640, ISBN: 9781510860964.

[72] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis*, 3rd ed. New York: Chapman and Hall/CRC, Jul. 2015, ISBN: 978-0-429-11307-9. DOI: 10.1201/b16018.

[73] S. He, Y. Li, Y. Feng, *et al.*, "Learning to predict the cosmological structure formation," en, *Proceedings of the National Academy of Sciences*, vol. 116, no. 28, pp. 13 825–13 832, Jul. 2019, ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1821458116. [Online]. Available: http://www.pnas.org/lookup/doi/10.1073/pnas.1821458116 (visited on 12/02/2021).

[74] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," *arXiv:1312.6114 [cs, stat]*, May 2014, arXiv: 1312.6114. [Online]. Available: http://arxiv.org/abs/1312.6114 (visited on 12/02/2021).

[75] S.-Y. Li, Y.-L. Li, and T.-J. Zhang, "Model comparison of dark energy models using deep network," en, *Research in Astronomy and Astrophysics*, vol. 19, no. 9, p. 137, Sep. 2019, Publisher: IOP Publishing, ISSN: 1674-4527. DOI: 10.1088/1674-4527/19/9/137. [Online]. Available: https://doi.org/10.1088/1674-4527/19/9/137 (visited on 12/02/2021).

[76] M. Bartelmann and P. Schneider, "Weak gravitational lensing," en, *Physics Reports*, vol. 340, no. 4, pp. 291–472, Jan. 2001, ISSN: 0370-1573. DOI: 10.1016/S0370-1573(00)00082-X. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S037015730000082X (visited on 12/02/2021).

[77] A. J. Ross, F. Beutler, C.-H. Chuang, *et al.*, "The clustering of galaxies in the completed SDSS-III Baryon Oscillation Spectroscopic Survey: observational systematics and baryon acoustic oscillations in the correlation function," *Monthly Notices of the Royal Astronomical Society*, vol. 464, no. 1, pp. 1168–1191, Sep. 2016, ISSN: 0035-8711. DOI: 10.1093/mnras/stw2372. eprint: https://academic.oup.com/mnras/article-pdf/464/1/1168/18517623/stw2372.pdf. [Online]. Available: https://doi.org/10.1093/mnras/stw2372.

[78] D. Ribli, B. Á. Pataki, and I. Csabai, "An improved cosmological parameter inference scheme motivated by deep learning," en, *Nature Astronomy*, vol. 3, no. 1, pp. 93–98, Jan. 2019, Number: 1 Publisher: Nature Publishing Group, ISSN: 2397-3366. DOI: 10.1038/s41550-018-0596-8. [Online]. Available: https://www.nature.com/articles/s41550-018-0596-8 (visited on 07/03/2022).

[79] T.-X. Mao, J. Wang, B. Li, *et al.*, "Baryon acoustic oscillations reconstruction using convolutional neural networks," *arXiv:2002.10218 [astro-ph]*, Feb. 2020, arXiv: 2002.10218. [Online]. Available: http://arxiv.org/abs/2002.10218 (visited on 11/17/2020).

[80] S. Ravanbakhsh, J. Oliva, S. Fromenteau, *et al.*, "Estimating cosmological parameters from the dark matter distribution," in *Proceedings of The 33rd International Conference on Machine Learning*, M. F. Balcan and K. Q. Weinberger, Eds., ser. Proceedings of Machine Learning Research, vol. 48, New York, New York, USA: PMLR, 20–22 Jun 2016, pp. 2407–2416. [Online]. Available: https://proceedings.mlr.press/v48/ravanbakhshb16.html.

[81] C. Modi, Y. Feng, and U. Seljak, "Cosmological reconstruction from galaxy light: Neural network based light-matter connection," *Journal of Cosmology and Astroparticle Physics*, vol. 2018, no. 10, pp. 028–028, Oct. 2018. DOI: 10.1088/1475-7516/2018/10/028. [Online]. Available: https://doi.org/10.1088/1475-7516/2018/10/028.

[82] K. Han, Y. Wang, J. Guo, Y. Tang, and E. Wu, *Vision gnn: An image is worth graph of nodes*, 2022. DOI: 10.48550/ARXIV.2206.00272. [Online]. Available: https://arxiv.org/abs/2206.00272.

[83] J. Fluri, T. Kacprzak, A. Lucchi, *et al.*, "Cosmological constraints with deep learning from kids-450 weak lensing maps," *Phys. Rev. D*, vol. 100, p. 063 514, 6 Sep. 2019. DOI: 10.1103/PhysRevD.100.063514. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevD.100.063514.

[84] "Illustris - about." (), [Online]. Available: https://www.illustris-project.org/about/ (visited on 01/17/2023).

[85] Norma Araceli Nava Moreno, "Simulaciones a 1.1/1.4/2.0 mm de la estructura a gran escala del universo a estudiar con TolTEC," Accepted: 2019-11-07T14:43:46Z, Ph.D. dissertation, Nov. 7, 2019. [Online]. Available: http://inaoe.repositorioinstitucional.mx/jspui/handle/1009/1904 (visited on 01/17/2023).

[86] V. Springel, "The cosmological simulation code gadget-2," *Monthly Notices of the Royal Astronomical Society*, vol. 364, no. 4, pp. 1105–1134, Dec. 2005, ISSN: 0035-8711. DOI: 10.1111/j.1365-2966.2005.09655.x. eprint: https://academic.oup.com/mnras/article-pdf/364/4/1105/18657201/364-4-1105.pdf. [Online]. Available: https://doi.org/10.1111/j.1365-2966.2005.09655.x.

[87] A. Radford, L. Metz, and S. Chintala, "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks," *arXiv:1511.06434 [cs]*, Jan. 2016, arXiv: 1511.06434. [Online]. Available: http://arxiv.org/abs/1511.06434 (visited on 11/30/2021).

[88] A. Béres, *Keras documentation: Data-efficient GANs with Adaptive Discriminator Augmentation*, en. [Online]. Available: https://keras.io/examples/generative/gan_ada/ (visited on 11/28/2021).

[89] M. Waskom, "Seaborn: Statistical data visualization," *Journal of Open Source Software*, vol. 6, no. 60, p. 3021, Apr. 2021, ISSN: 2475-9066. DOI: 10.21105/joss.03021. [Online]. Available: https://joss.theoj.org/papers/10.21105/joss.03021 (visited on 12/03/2021).