High Resolution Synthetic Galaxy Image Generation using Deep Learning

Manop Chugh 60090016
Supavich Wongwachiraporn 60090038

Bachelor of Engineering in Software Engineering
Faculty of Enginerring
King Mongkut's Institute of Technology Ladkrabang
Academic Year 2020

Software Project - Academic Year 2020

Bachelor of Engineering in Software Engineering
Faculty of Engineering
King Mongkut's Institute of Technology Ladkrabang

Title: High Resolution Synthetic Galaxy Image Generation using Deep Learning
Authors:

1. Manop Chugh Student ID: 60090016
2. Suphavich Wongwachiraporn Student ID: 60090038

Approved for Submission

.....

(Asst.Prof.Dr. Isara Anantavrasilp)
Advisor

Date ...9.../...6.../...2021

Acknowledgments

This research is not possible without the help of Assistant Prof. Ukrit Watchareeruethai and Assistant Prof. Isara Anantavrasilp, our advisors, who give useful and practical advice for the efficient development of our research, comments and feedbacks that we can improve upon and provides us with many resources that has been tremendously beneficial for our research.. Secondly, our fellow colleagues have been generously providing useful feedback, comments, and criticizing us which ultimately contribute to the progress of this project.

Abstract

In this modern age, dark energy is still a topic of mystery in the field of astrophysics and cosmology. The next generation of cosmological surveys and research depends upon considerable amounts of astrophysical data such as high resolution images of galaxies. To conduct future research in the field of dark energy, an accurate measurement and shape of distant galaxies are needed. However, high resolution galaxy images are difficult to obtain or generate, we have seen previous work using GANs to generate galaxy images with a maximum resolution of 128×128, which does not guarantee the visibility of the structure of the galaxy. Here we show our approach of generating high resolution galaxy images using a modified version of SAGAN [8] combined with a super resolution network which we introduce as Joint Upscaling Adversarial Network or JUAN.

Table of contents

List of figures

List of tables

Chapter 1
Introduction

1.1 Problem Description

Since the late 1950s and the beginning of the space age, we have discovered and confirmed many theories relating to deep space, such as black holes, galaxies, or exoplanets, just to name a few. We have also confirmed and have robust evidence that the universe is expanding and the expansion is accelerating [1]. The only explanation of the expansion of the universe is the presence of mysterious forces undetectable to mankind, also known as dark energy. This suggests that the majority of the energy content in the universe is dark energy [2]. It is no mystery that we still lack knowledge and understanding of what dark energy actually is, which provides one of the main motivations behind the next generation of cosmological surveys.

Cosmological surveys such as surveys related to dark matter and dark energy need huge amounts of astrophysical data such as images of astronomical objects like galaxies and stars to simulate the astronomical events and unfold the mystery of dark energy and dark matter. One way of learning about dark energy is by using the property of weak gravitational lensing that is caused by galaxy images that are slightly stretched or sheared when observing from telescopes like the Hubble telescope. These warped or stretched galaxy images indicate the presence of dark energy or dark matter which then can be mapped where the dark matter is spread out around the galaxies. It is, however, very difficult to collect high-resolution galaxy images to simulate these cosmological surveys very well.

There have been several approaches [5, 6] to generate large amounts of galaxy images using Generative Adversarial Networks (GANs) [4] and Variational Autoencoders (VAEs) [3] to use in cosmological surveys, however, the results of these approaches are not directly applicable to the said cosmological surveys or further dark energy research because of their low resolution and less visibility of the features such

as spiral arms in a spiral galaxy.

We intend to solve the problem of low-resolution galaxy image generation by experimenting with the use of newer architectures of GANs (like Self-attention GAN) [8] or creating a tailor-made GAN based on some previous architecture of GANs for the specific task of galaxy image generation. We also propose to take the output from the chosen GAN architecture and input it into a super-resolution algorithm.

1.2 Objective

1. The objective of this research is to be able to generate large amounts of galaxy images with variety and which is accurate and the generated images are classifiable to their respective structures in the Hubble Galaxy Classification [7] as shown in fig 1.1 by using SAGAN [8] because of its ability to keep important structures in an image such as the structure of the galaxy.

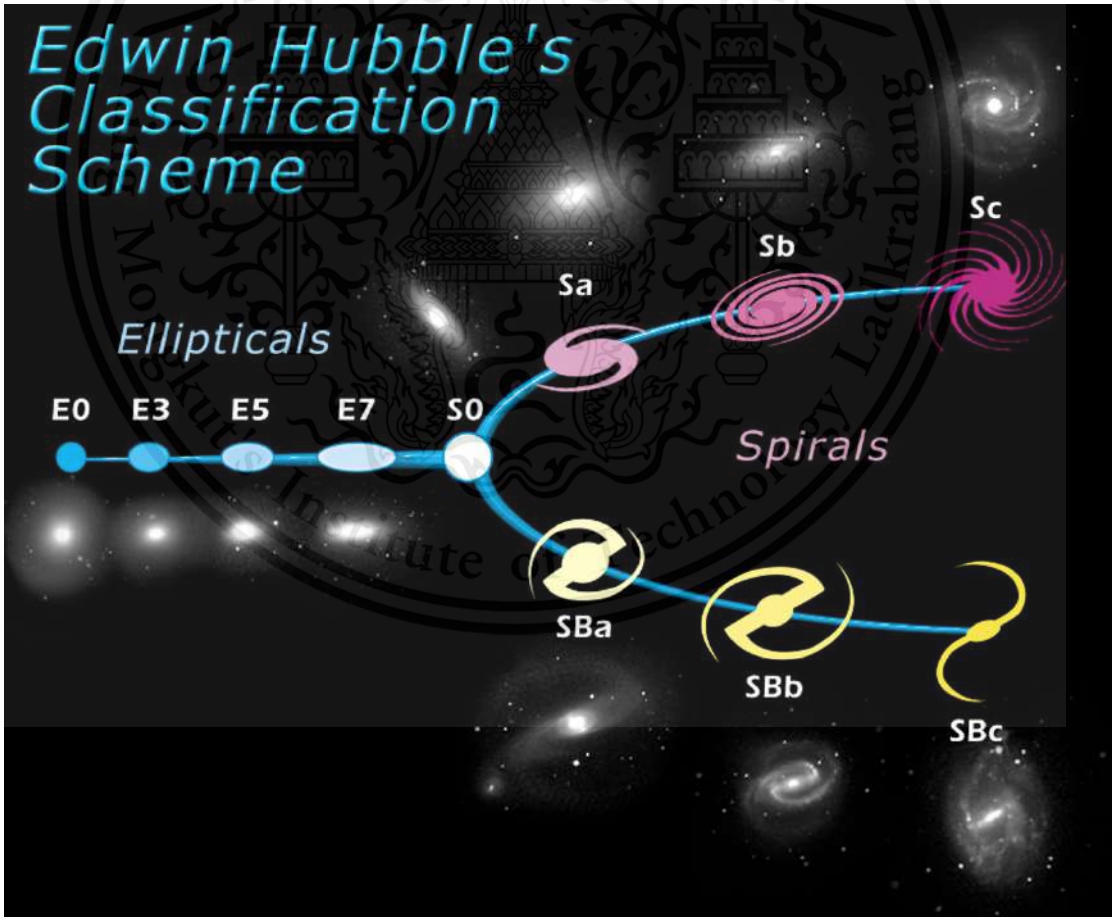

Figure 1.1: Hubble Galaxy Classification, (NASA ESA)

2. The features of the generated galaxy images must be clearly visible such as the spiral arm with a resolution of at least 256×256

1.3 Thesis Structure

The remaining of this thesis consists of six more chapters, as follows:

- Chapter 2 proposes the literature survey, a various software that are relevant to this project, and comparison.
- Chapter 3 explains the knowledge and technology necessary for the reader to understand the thesis.
- Chapter 4 presents our approach and contribution to full fill the objective of this research.
- Chapter 5 explains the concepts, tools, and techniques that are used in developing the project.
- Chapter 6 refers to the experimental results of our approach and how well it completes the objectives of this research.
- Chapter 7 Concludes our research and summarizes the steps and the result of our research.

Chapter 2
Related Works

2.1 Enabling Dark Energy Science with Deep Generative Models of Galaxy Images

Ravanbakhsh et al. [5] talked about the importance of generating images of galaxies by pointing out how difficult it is to obtain images of a galaxy that has enough resolution to be used for instrument calibration or simulation, which will be used for dark energy surveys. Their approach is by using Variational Autoencoders (VAEs) [3] and Conditional GANs (C-GAN) [19] which is trained from the GalaxyZoo dataset [20].

2.1.1 Importance of generating images of galaxies

Ravanbakhsh et al. [5] explain in the paper why generating galaxy images are needed in cosmological surveys. In the field of astrophysics, we still lack an understanding of what dark energy is, which is one of the main motivations behind the upcoming cosmological surveys such as LSST [22]. These kinds of cosmological surveys are designed to give us some lead or to direct us to the right path of understanding the nature of dark energy by probing the Universe through the weak gravitational lensing effect. On the big scale, these lensing effects cause very small but still visible deformations of background galaxy images, which might appear slightly sheared. This property of weak gravitational lensing may provide a way to statistically map the matter distribution in the Universe. To measure these lensing signals, upcoming cosmological surveys will need to image and measure the shapes of millions or billions of galaxies to drive down statistical errors compared to the current generation of surveys.

The quality of the results of these cosmological surveys relies greatly on the accuracy of the shape measurement algorithm tasked with estimating the ellipticity of the galaxies in the survey. Should any unaccounted measurement be biased, the results of

these surveys would be heavily impacted. Therefore, to account for these biases and calibrate these measurements, cosmological surveys in the future will heavily rely on image simulations, closely mimicking real observations.

Obtaining or producing these image simulations is very challenging and requires high-resolution astrophysical images, such as galaxy images. Therefore, in this work, Ravanbakhsh et al. [5] proposed an alternative to the expensive acquisition of high-quality calibration data such as galaxy images using deep generative models, specifically Variational Autoencoders (VAEs) [3] and Conditional GANs (C-GAN) [19].

2.1.2 Results of using VAE's to generate images of galaxies

Ravanbakhsh et al. [5] show us that the images of galaxies generated using VAE's are low in resolution and may not show the main features of galaxies clearly such as spiral arms.

2.2 Forging New Worlds: High-Resolution Synthetic Galaxies with Chained Generative Adversarial Networks

Fussell et al. [6] talk about how the interest in using machine learning for tasks such as galaxy processing, classification, and segmentation has become popular due to the availability and growth of larger galaxy datasets. As more of these approaches become more complex and more interest is seen in the task of galaxy processing using machine learning, more data is required to do such tasks.

Fussel et al. [6] propose that GANs [4], a type of unsupervised machine learning model, should be used to generate images of galaxies to increase the data availability.

Fussel et al. [6] propose that using DC-GAN [23] works well for galaxy image generation since using convolution layers will decrease the number of parameters per layer. However, the problem with DC-GAN is it does not perform well when generating images larger than 64×64, hence Fussel et al. propose the use of StackGAN [24] for super-resolution, to increase the resolution of the output of the DC-GAN which is 64×64 to 128×128.

The model proposed by Fussel et al. [6] is able to produce galaxy images with

some features visible such as the spiral arms and bright center of the galaxy, however, the results are inconsistent and may not always show these features clearly. The results of the model proposed by Fussel et al. [6] were using DCGAN [23] and StackGAN [24] for super-resolution, however, the limitation of the image size is still a problem, the output of the model proposed by Fussel et al. [6] is 128×128.

2.3 BigGANs and Not-so-BigGAN

From the BigGANs model by Andrew et al. [11], GANs model can greatly benefit from upscaling the size of the training batch and the depth of the model itself, which can generate an image of resolution up to 512×512 but due to the larger batch size and having the total number of parameters skyrocketed can be detrimental towards a limited capability hardware environment.

Not-so-BigGan by Seungwook et al. [13] on the other hand approaches this problem differently, by operating mainly on the frequency domain and different model architecture than BigGAN. This allows not-so-BigGAN to perform on a level similar to that of the state-of-the-art models with a relatively lower computational budget and time. This recent progress towards a higher fidelity image generation model will help our model to have a quality of galaxy images like never before.

Chapter 3
Background Knowledge

In this chapter, relevant background knowledge is described in order to get a better understanding of the research, which includes topics in astrophysics and deep learning, hence this chapter will be divided into two main parts.

3.1 Astrophysics

3.1.1 Galaxy

A galaxy is a gravitationally bound system of stars, stellar remnants, interstellar gas, dust, and dark matter [9, 10]. Earth is in a star system called the Solar system which among many other star systems are in a galaxy, which in our case is called the Milky Way galaxy, a type of spiral galaxy, we will elaborate on the types and classifications of galaxy further down this chapter. Here we show an example of our neighbour galaxy called Andromeda in Fig. 3.1

Figure 3.1: M31, the Andromeda Galaxy - Image By Adam Evans

There are three main types of galaxies that can be defined, with each type having a distinct feature that is easily recognizable.

Elliptical

The Hubble classification Scheme rates elliptical galaxies on the basis of their ellipticity, ranging from E0, being nearly spherical, up to E7, which is highly elongated as shown in Fig. 1.1. These galaxies have an ellipsoidal profile, which maintains the elliptical experience regardless of the view angle. The appearance of elliptical galaxies are smooth ellipsoids without any distinct structure as compared to spiral galaxies which have a clear structure which is the spiral arms.

The largest galaxies are ellipticals as they can grow to enormous sizes compared to spiral galaxies. Most stars in an elliptical galaxy are on average older stars than the stars in spiral galaxies. Many elliptical galaxies form due to the interaction of galaxies, for example, collision and merging of spiral galaxies [12]. An example of the giant elliptical galaxy ESO 325-G004 is shown in Fig. 3.2.

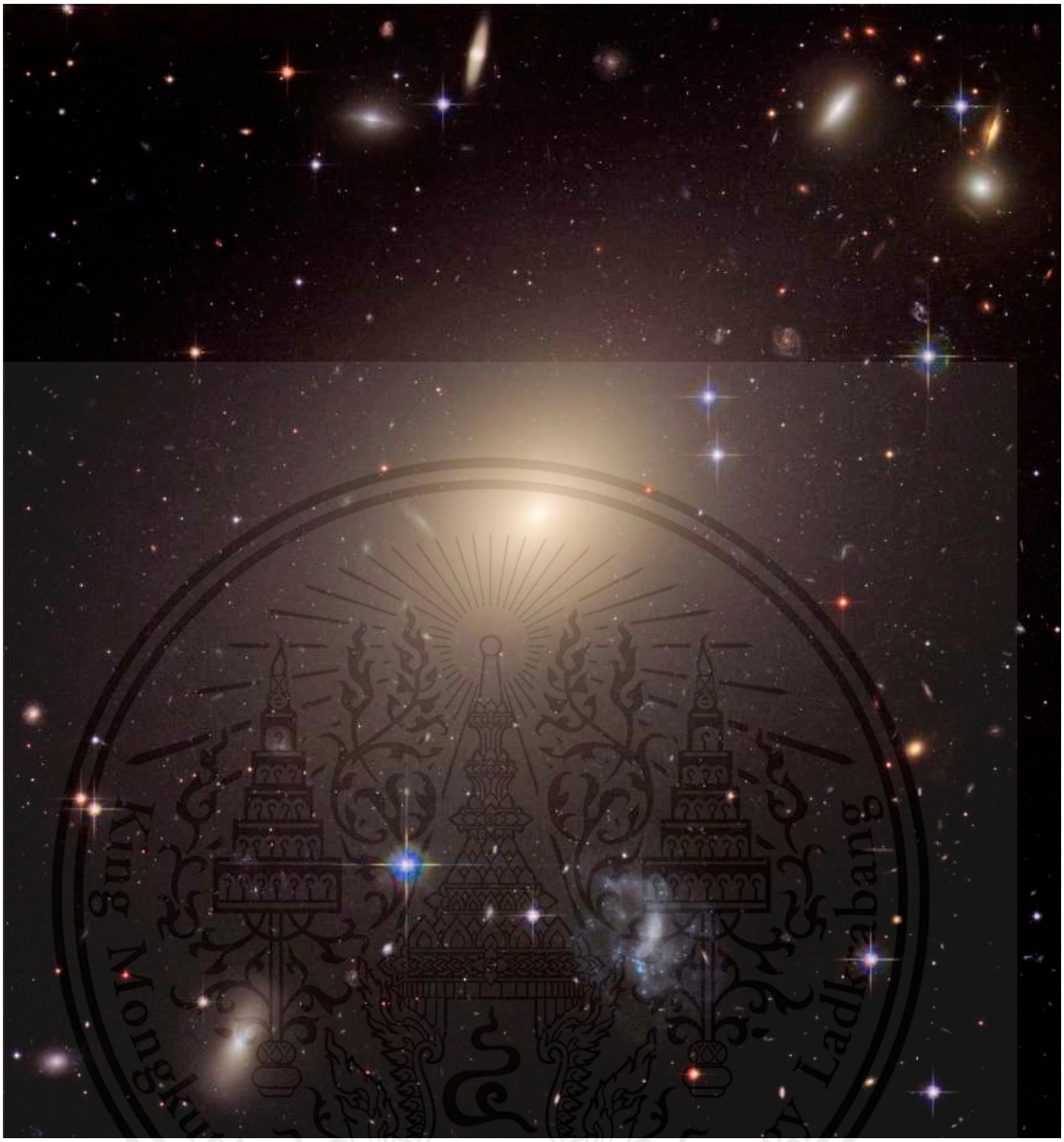

Figure 3.2: The giant elliptical galaxy ESO 325-G004 (NASA, ESA, and The Hubble Heritage Team (STScI/AURA); J. Blakeslee (Washington State University))

Spiral

Spiral galaxies resemble a disc of spiraling pinwheels. The majority of the stars, gas, and dust lie on a plane. The majority of mass in a spiral galaxy is in the form of dark matter, which is considered as the nonvisible part of the galaxy which is roughly in the form of a spherical halo, as demonstrated by the universal rotation curve concept [14].

Spiral galaxies consist of a rotating disk of stars and interstellar gas and dust, and a bulge of older stars towards the middle of the galaxy. The arms of a spiral galaxy

extend outward which forms the classic spiral look. In the Hubble classification scheme, spiral galaxies are listed as type S, followed by a letter (a, b, or c), which indicates the degree of tightness of the spiral arms and the size of the central bulge. The Sa galaxy has tightly wound, poorly defined arms and possesses a relatively large core region. At the other extreme, an Sc galaxy has open, well-defined arms and a small core region [7]. An example of the famous spiral galaxy called the Pinwheel Galaxy is shown in Fig. 3.3.

Figure 3.3: The Pinwheel Galaxy (ESA/Hubble)

Irregular

Irregular galaxies are galaxies that are not classifiable into either elliptical or spiral. Irregular galaxies can be found in any shape with no consistent structure or properties. Irregular galaxies are usually a newly forming galaxy or galaxies that are merging, it could also be a sub-galaxy of another galaxy. An example of an Irregular galaxy. An example of an Irregular galaxy NGC 1427A can be seen in Fig 3.4.

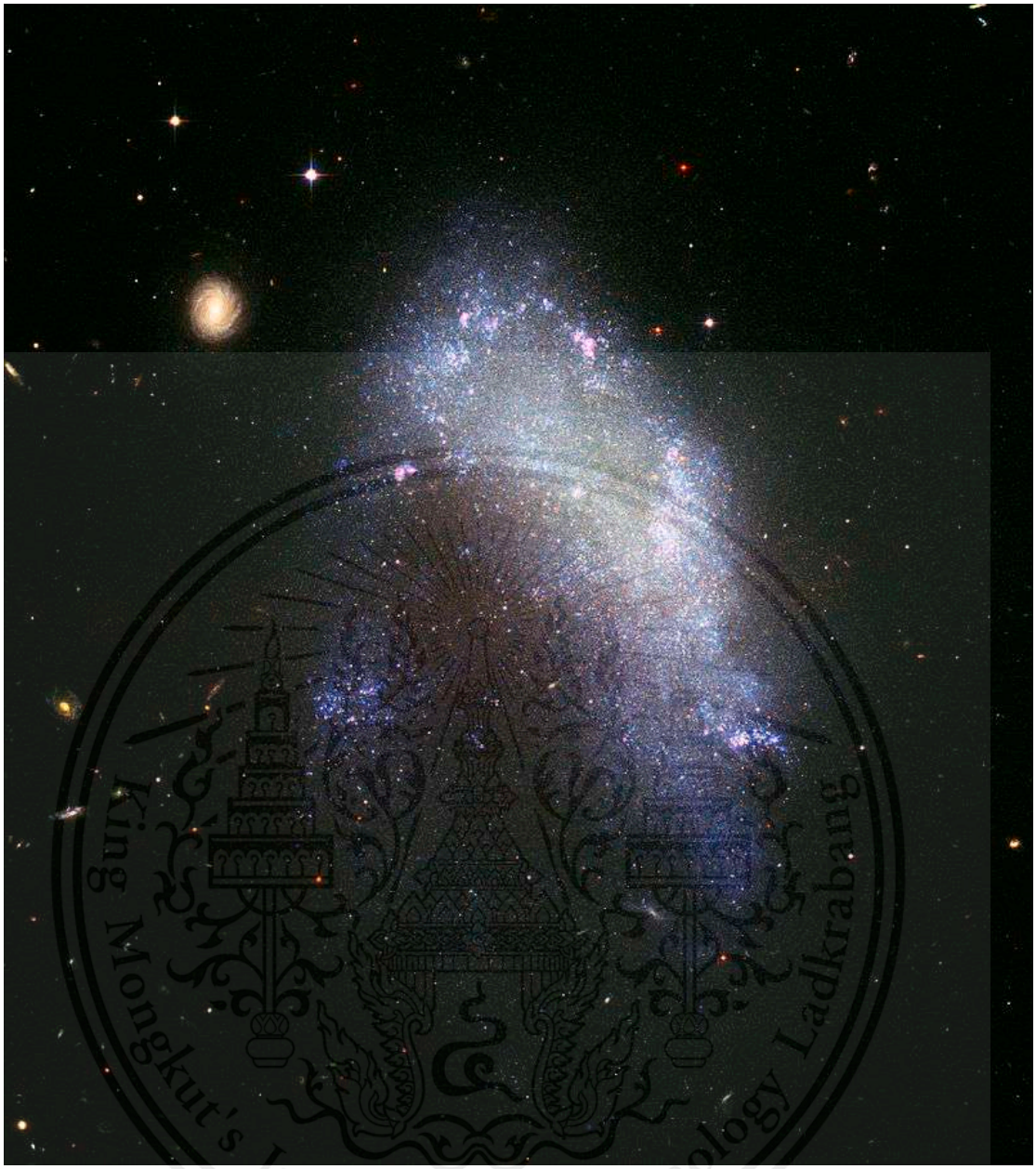

Figure 3.4: NGC 1427A (NASA, ESA, and The Hubble Heritage Team (STScI/AURA))

3.1.2 Dark Energy

In astrophysics, cosmology, and astronomy, dark energy is an unknown form of energy that affects the universe on the largest scales. The first observational evidence for its existence came from supernovae measurements, which showed that the universe does not expand at a constant rate; rather, the expansion of the universe is accelerating [10].

The volume of dark energy in the universe is thought to be over 70% as takes immense force to push all interstellar masses such as galaxies, stars, gas, and dust out expanding and an accelerated rate of expansion. Figure 3.5 shows snapshots from a simulation by astrophysicist Volker Springel of the Max Planck Institute in Germany. It represents the growth of cosmic structure (galaxies and voids) when the universe was 0.9 billion, 3.2 billion, and 13.7 billion years old (now)

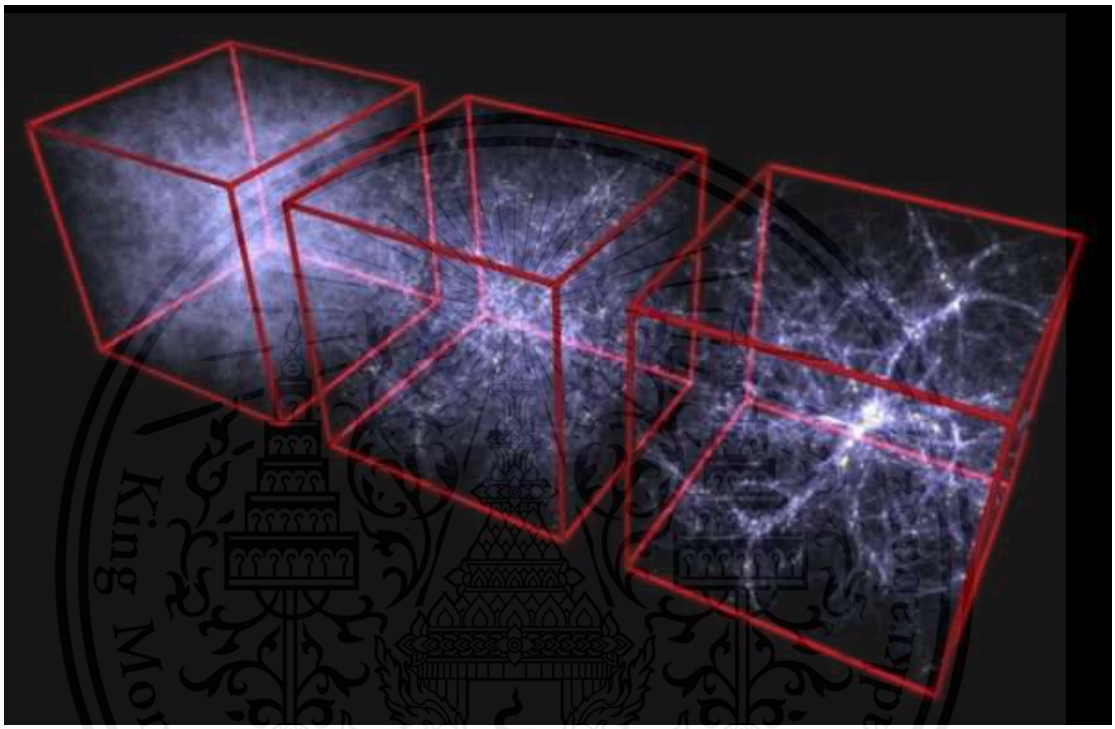

Figure 3.5: Simulation showing structure (galaxies and voids) when the universe was 0.9 billion, 3.2 billion and 13.7 billion years old (now) . Image via Volker Springel/ MPE/Kavli Foundation. (STScI/AURA))

The closest candidate we have to dark energy is the cosmological constant, which proves that the influence of dark energy is repulsive or anti gravity. This can be seen from Albert Einstein's General Theory of Relativity [31].

3.2 Relativity, the Special and General Theory

3.2.1 Special Relativity

Theorized in 1905 by Albert Einstein and disproving Newton's theory that time is absolute. Einstein argues and proves that time in fact, is not absolute.

Basic idea

The first thing that came to the mind of young Albert Einstein is, can you tell if something is in motion in a space without any other object? Imagine a real empty void with ball accelerating infinitely, there is no air, no air resistance, no stars, no galaxies, nothing to compare to, how can you tell if the ball is moving or not? It turns out you can't because movement is a relative motion.

The basic idea of relativity is that two different observers who are in motion relative to each other must agree on the laws of physics. When two different observers are in relative motion, they are said to be in different reference frames, and when their relative velocity is constant, those reference frames are said to be an inertial frames of reference.

A main fact to note is: the speed of light is constant for all observers and if there is anything that's changing it's time.

Assume an observer at rest with respect to a frame S and the said observer is standing at the origin of the coordinate system (x,y,z), and that an observer at rest with respect to frame S' is standing at the origin of a coordinate system (x',y',z'). If the observer in frame S sees the origin of the frame S' coordinates moving to the right with a constant velocity V, then the two reference frames S and S' are said to be in standard configuration.

Time dilation

Special relativity indicates that, for an observer in an inertial frame of reference, a clock that is moving relative to them will be measured to tick slower than a clock that is at rest in their frame of reference. This case is sometimes called special relativistic time dilation.

3.2.2 General Relativity

Space and time

$$R_{\mu\nu} - \frac{1}{2}R\, g_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu} \qquad (3.1)$$

Where:
$R_{\mu\nu}$ is the Ricci curvature tensor.
$g_{\mu\nu}$ is the metric tensor.
R is the scalar curvature.
Λ is the Cosmological constant.
c is the speed of light in vacuum.
G Newton's Gravitational constant.
$T_{\mu\nu}$ is the stress energy tensor.

Equation 3.1 is the Einstein field equations which relate the geometry of space-time to the distribution of matter within it [31], or in simple words, the Einstein field equations can explain the curvature, stretching or expansion of the fabric of space and time. The Einstein field equations were first published by Einstein in 1915 in the form of a tensor equation, since the tensor variables are used to represent the multi dimensional model and curvature of space-time.

In the scope of this research, since the primary objective is to generate high resolution images of galaxies with visible features to simulate the prominent distribution of dark energy, we will extract the variable that is most important to us, which is the Λ variable. This lambda variable is known as the cosmological constant, and the other variables such as the Ricci curvature tensor, the metric tensor and the scalar curvature are not constants. To put it simply, the other values which represent the curvature or expansion of space will expand further, and the expansion will accelerate, because the cosmological constant is not changing while the space is expanding, which means the cosmological constant is relatively getting more and more concentrated, as space gets less dense and far apart. Physicists have concluded that the Lambda variable or the cosmological constant in the Einstein Field Equation best represents the model of dark energy as we understand right now [31].

Gravitational lensing

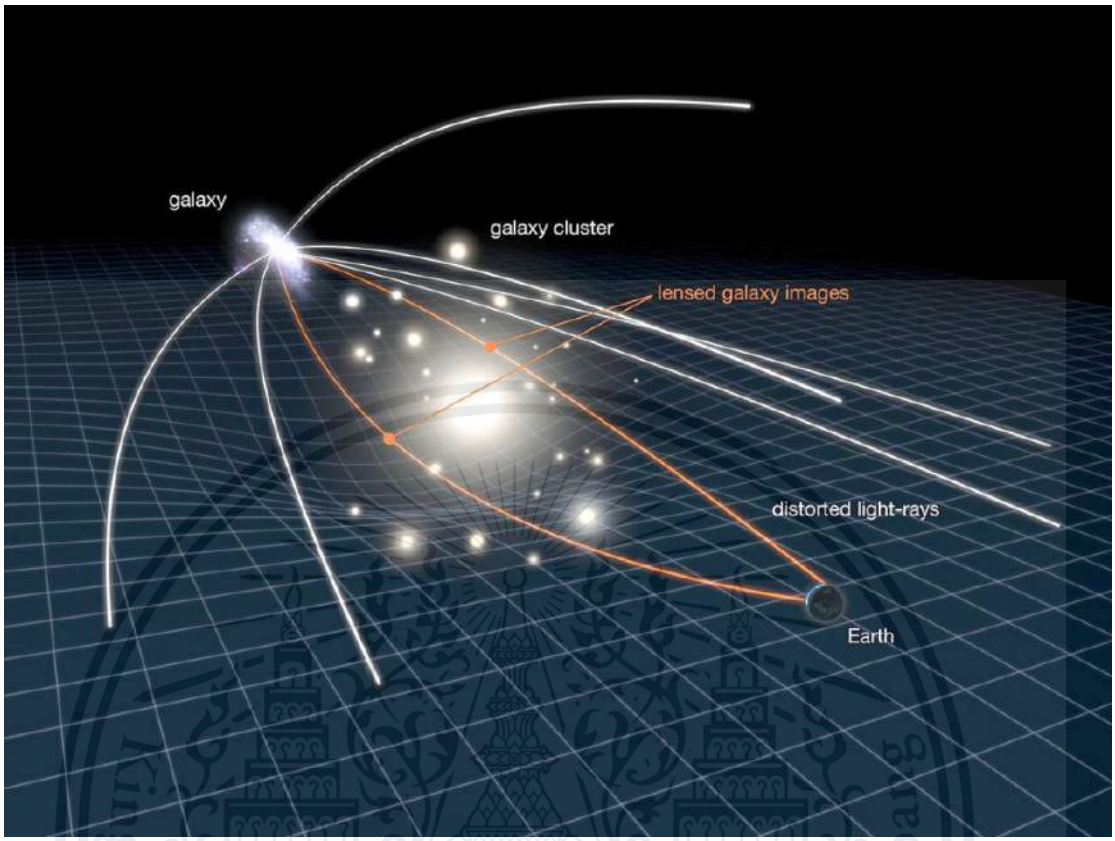

Figure 3.6: Visualization of gravitational lensing . Image via CFHTLenS [https://www.cfhtlens.org/public/what-gravitational-lensing]

Fig 3.6 shows a visualization of gravitational lensing, when there is an object with a relatively large mass and the object behind, in this case, a spiral galaxy. The light from the galaxy travels in a straight path until the disturbance in the space-time from the massive object which curves the light and it could appear stretched, sheared or oddly curved from the original image, and not only can normal matter cause this, but also dark energy and dark matter, although it would be a weak gravitational lensing in the case of dark energy [31].

Figure 3.7: Visualization of gravitational lensing . Image via BBC Sky at Night Magazine [https://www.skyatnightmagazine.com/space-science/a-guide-to-gravitational-lensing/]

Fig 3.7 shows a real example of gravitational lensing easily detectable by the slightly curved lines which are supposed to be a normal galaxy.

3.3 Deep Learning

We will explain the basics of deep learning and mainly, Generative Adversarial Networks (GAN) [4] which is the core part of this research.

3.3.1 Neural Network

A neural network, or more precisely, an artificial neural network (ANN), is a network or circuit of neurons (nodes). An artificial neural network is composed of nodes and connections which are modeled as weights. All inputs are modified by weight and summed which is known as a linear combination. The output is then passed to an activation function or a step function to control the amplitude of the output, for example, the output could be in between 0 and 1 or -1 and 1. A perceptron can be considered as a single-layer neural network, an example of a perceptron is shown in Fig 3.8.

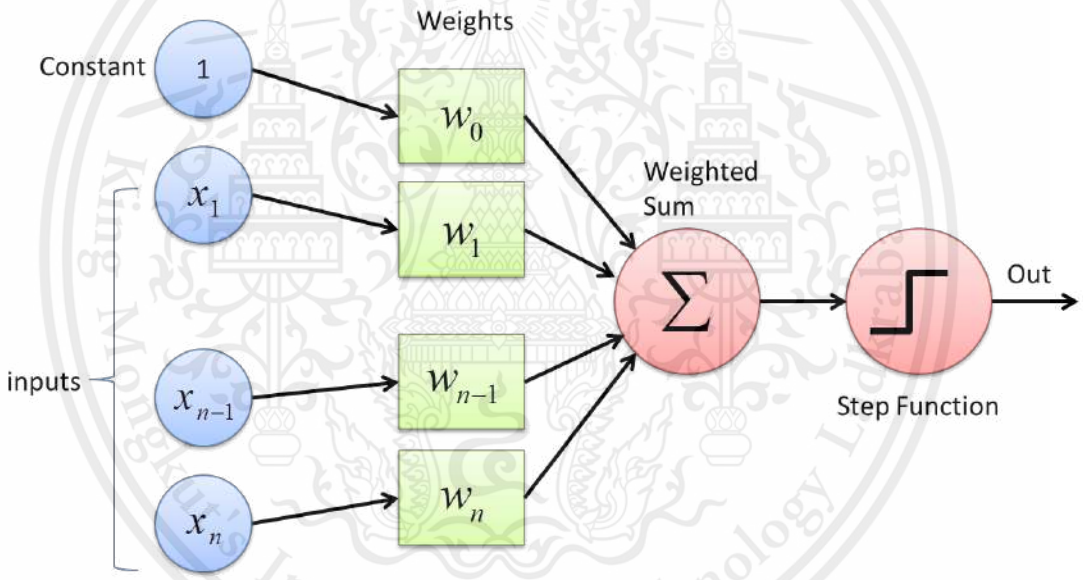

Figure 3.8: A perceptron
(https://deepai.org/machine-learning-glossary-and-terms/perceptron)

Fig. 3.8 shows that the process begins by taking all the input values and multiplying them by their weights. Then, all of these multiplied values are added together to create the weighted sum. The weighted sum is then applied to the activation function, producing the perceptron's output.

3.3.2 Backpropagation

To train our neural network to achieve better-classifying accuracy, A training method called Backpropagation is used. By applying a cost function to an output layer, we can determine the distance between the current output and the correct classifying output. Now that the neural network has a baseline on what the correct output should be, the weights in the network will slowly adjust from the layer right before the output layer back to the very first input layer so that the same input will yield an output value that is increasingly similar to the correct set of outputs in each training iteration, narrowing down the distant measured by a loss function [15].

3.3.3 Deep Learning

Deep learning [16, 17, 18] is simply the process of using multilayered artificial neural networks and train them using learning algorithms like backpropagation. Some examples of deep learning architectures are:

- Deep Neural Networks
- Recurrent Neural Networks
- Deep Belief Networks
- Convolution Neural Networks
- Autoencoders

The image in Fig. 3.9 shows the difference between a normal neural network model and a deep neural network model.

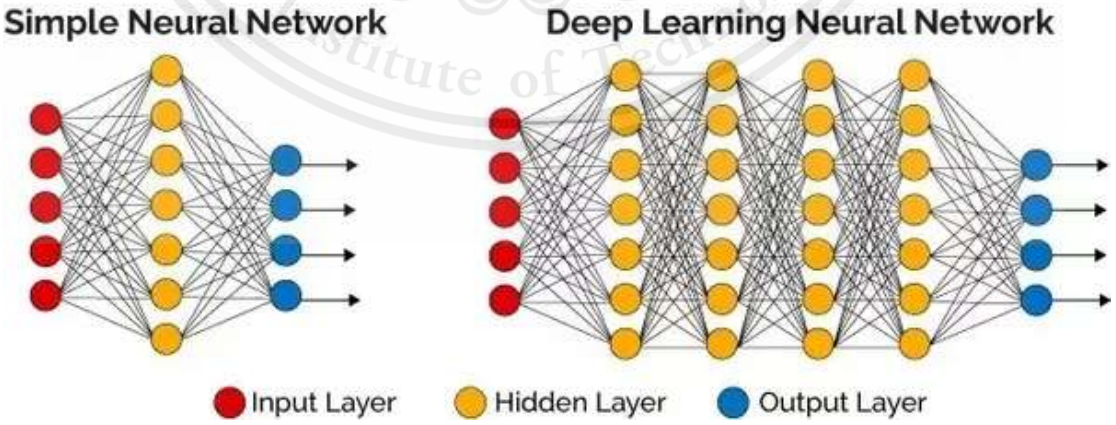

Figure 3.9: A comparison between normal neural network and a deep neural network (https://thedatascientist.com/what-deep-learning-is-and-isnt/)

3.4 Introduction to Generative Models

3.4.1 Definition of Model

In science and engineering, we describe a system using mathematical concepts and language. The description is called as a mathematical model, and the process of developing a mathematical model is mathematical modeling. Especially, in the context of machine learning, we explain a target model by a map f from an input x to an output y

$$f : x \to y$$

Therefore, the purpose of model training is obtaining the map f from the training data. In the case of unsupervised learning, we use the dataset of inputs $s(n) = d1, d2, ..., dN$ as the training data, and create the model f. In supervised learning, we use the dataset of inputs and their outputs $s(n) = (d1, c1), ..., (dN, cN)$. As a simple example, let's consider a supervised learning problem such as classifying images into dogs or cats. Then, the training dataset consists of input images $d1, d2, ..., dN$ and their labels c1=cat,c2=dog,...,cN=cat.

3.4.2 What is the Generative Model?

The generative model models the probability distribution $p : s \to p(s)$ which generate the training data s. One of the simplest ways to think about a generative model is that a generative model models the probability distribution p with the map f. We assign each x and y of $f : x \to y$ as follows.

- x: the training data s
- y: the likelihood of generating the training data s

3.4.3 Autoencoders

An autoencoder [21] is an unsupervised artificial neural network that learns how to efficiently compress and encode data and then learns how to reconstruct the data back from the reduced encoded representation to a representation that is as close to the original input as possible. To put it simply, an autoencoder is a neural network that learns to copy its input to its output. It has a hidden layer that describes the code used to represent the input, and it is constituted by two main parts: an encoder that maps the input into the code, and a decoder that maps the code to a reconstruction of the original input.

If an autoencoder can perform the copying task perfectly, that would mean it is just duplicating the signal exactly as it was input, and this is why autoencoders are restricted in ways that force them to reconstruct the input approximately, which eventually preserves only the most relevant features of the data in the copied version.

The main purpose of an autoencoder is to learn in an unsupervised manner, the reconstructed data is a new representation of which is stripped to reveal the main features. This representation can prove to be useful for various applications, such as clustering, generative models, or dimensionality reduction.

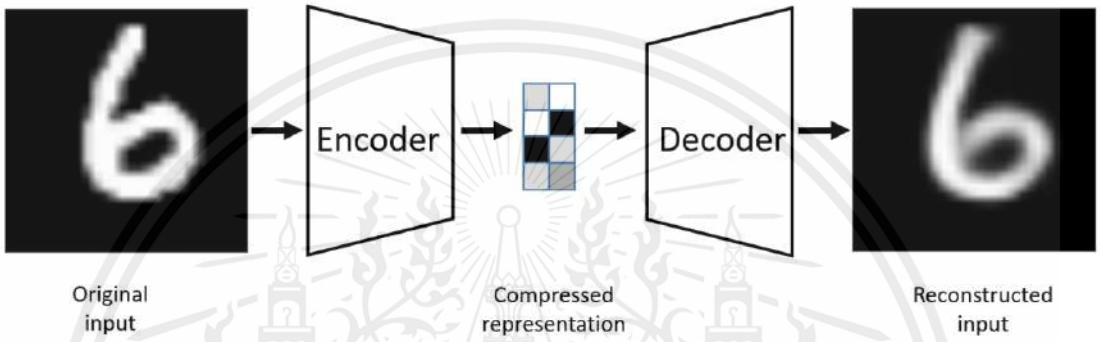

Figure 3.10: Autoencoder [21].

Figure 3.10 shows an example of an autoencoder, how the input image is encoded to a compressed representation and then decoded.

An autoencoder can be represented as a problem to learn the functions $A : \mathbb{R}^n \rightarrow \mathbb{R}^p$ (encoder) and $B : \mathbb{R}^p \rightarrow \mathbb{R}^n$ (decoder) satisfy

$$argmin_{A,B} E[\Delta(x, B \circ A(x))].$$

Where,E is the expectation over the distribution of x, Δ is the reconstruction loss function, and which measures the distance between the output of the decoder and the input.

3.4.4 Generative Adversarial Network

A generative adversarial network, which is first introduced by Goodfellow et al. in 2014 [4]. Generative adversarial network (GAN) is a deep learning model in which two multilayer perceptrons play a minimax game with the following objective.

$$\min_{G} \max_{D} V(D,G) = \mathbb{E}_{x \sim pdata(x)} [\log D(x)] + \mathbb{E}_{z \sim pz(z)} [\log(1 - D(G(z)))]$$

In GAN model, there exist two multilayer perceptrons which are ***D***iscriminator and ***G***enerator. ***G*** main purpose is to map an set of latent input noise into an output in order to recreate a data distributions $p_z(z)$ of a real $p_{data}(x)$. Meanwhile, ***D*** will learn to discriminate the two distribution of a $p_{data}(x)$, $D(x)$ and $p_z(z)$ that is generated by $G(z)$,$D(G(z))$.

In this minimax game, a cross-entropy log loss is used for both ***G*** and ***D***. Given that $logD(x)$ represent a likelihood of ***D*** discriminating a data in favor of $p_{data}(x)$ and $log(1 - D(G(z)))$ represent a likelihood of ***D*** discriminating over $G(z)$, ***D*** wants to maximize this equation and ***G*** wants to minimize the value of this function by its influence over $log(1-D(G(z)))$. During the GAN training process, ***G*** will learn to generate an image from a latent space, that is resembled closely to the real image data distribution $p_{data}(x)$ and ***D*** will be try to detect if the $G(z)$ belongs to the $p_{data}(x)$ or not hence they will both take turn to improve each other's ability [4]. GAN architecture is one of the most groundbreaking networks introduced in recent years and has been a staple in many image generating deep learning models since.

3.4.5 Deep Convolutional Generative Adversarial Network

In this section, we will introduce the model called DCGAN or Deep Convolutional Generative Adversarial Network proposed by Radford et al. DCGAN plays a pivotal role in the future of GANs in general, and the paper demonstrated that DCGAN is a strong candidate for unsupervised learning. As shown below in Fig3.11, it is a model using CNN (Convolutional Neural Network) as its name suggests.

The key points from the paper are the following [23]:

- The paper proposed an architectural constrain of DCGANs which could be incliDently trained, stable, and more likely to converge.
- the paper proved that DCGANs trained on a massive amount of images without labels, which could be a very competitive feature extractor, from both generators and discriminator (mostly from discriminator), that is used for other higher level supervised task, like the image classifier.
- The paper tried to peek how does the generator work and what kind of features and filters do they learn.

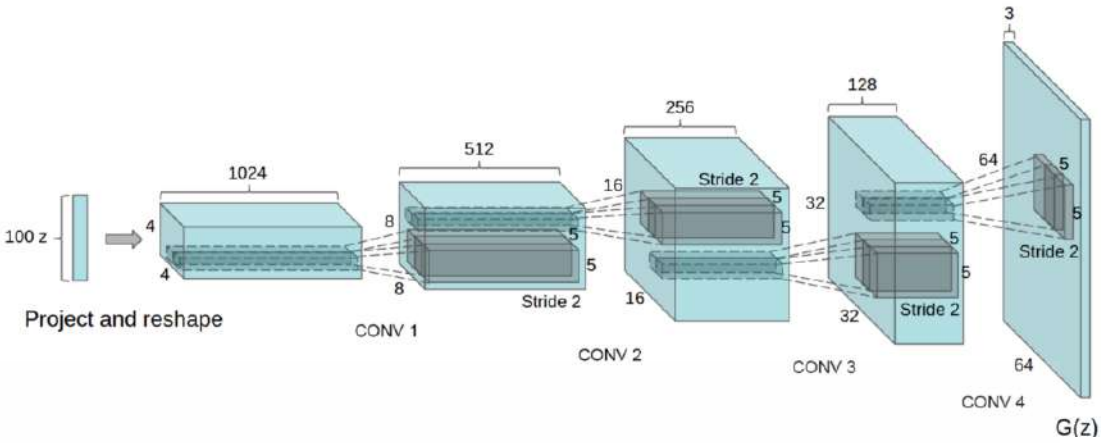

Figure 3.11: DCGAN generator used for LSUN scene modeling. A 100 dimensional uniform distribution Z is projected to a small spatial extent convolutional representation with many feature maps. A series of four fractionally-strided convolutions (in some recent papers, these are wrongly called deconvolutions) then convert this high level representation into a 64 × 64 pixel image. Notably, no fully connected or pooling layers are used [23].

- The paper shows that the generators have interesting vector arithmetic properties allowing for easy manipulation of many semantic qualities of the generated samples.

3.4.6 Self-Attention Generative Adversarial Networks

Attention Mechanism

To understand Self-Attention Generative Adversarial Networks, it is crucial to know the reason behind its main key feature, self-attention mechanism introduced by Cheng et al. [27].

Self-Attention mechanism is initially used in Natural Language Processing to identifies part of speech and how each words in a sentence related or refered to each other, as for example in figure 3.12 below.

It is clear that when self-attention mechanism is used, after iterating word by word in a sentence, it can pick up the correlation it has with other words at different activation level.

In a context of computer vision and image generation, a self-attention mechanism that does not rely on sequential data is proposed by Vaswani et al. [28]. This transformer-based self-attention module that utilizes key,query and value will captures in the information in an images based on their context. This allow SAGAN to adopt the self-attention module into its generator and discriminator.

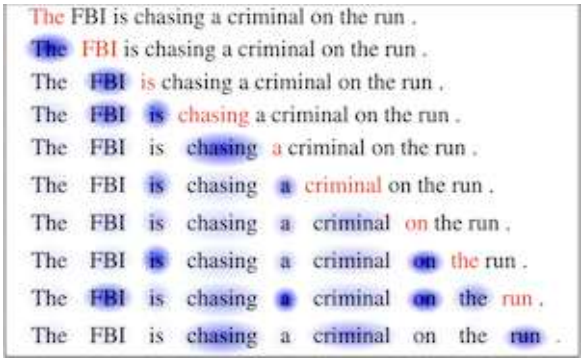

Figure 3.12: Self-Attention enables the context at which words correlate to what other words [27].

SAGAN

One of the limits of normal GAN, which can be summarized in that GAN excels at generating small fine details of a particular region of an image but performs relatively poorly in discovering long-range details of a different part of an image due to how the convolutional networks work. In order to generate a galaxy image, the overall structure of a galaxy is an integral part. However, to leverage against this problem caused by local neighbourhood convolutional operator in the networks, a new self-attention generative adversarial (SAGAN) networks are introduced by Han Zhang et al. in 2019 [8].

SAGAN achieves long-distance feature generated images by reinforcing self-attention in $\boldsymbol{G}$ and $\boldsymbol{D}$'s convolutional layer by performing an additional operation.

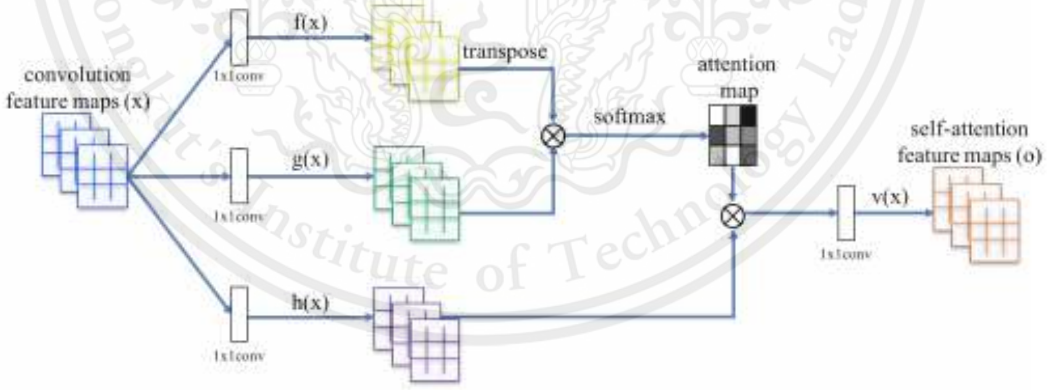

Figure 3.13: A self-attention operation diagram [8].

In the self-attention process of SAGAN, feature maps from the previous layer, x will undergo 3 additional convolution operations with three 1x1 filters which are f,g and h filters. $f(x)^\top$ and $g(x)$ will be multiplied to form an attention map that later will be, again, multiplied with $h(x)$ will yield a latent feature maps, as shown in Fig.**??**. The

final step is to multiplied the latent output from the previous step with filter v, resulting in an output feature maps of o. The entire self-attention process can be described in the following formula, where $W_y x_i$ means convolutional operator of any filter y with an input x_i, $\beta_{j,i}$ referred to the attention maps and the input is in the spatial dimension of $R^{M \times N}$ with output $o \in \{o1, o2, ..., o_j, ..., o_M\}$. The following equation denotes the self-attention module.

$$O_j = v\left(\sum_{i=1}^{N} \beta_{i,j} h(x_i)\right), h(x_i) = W_h x_i, v(x_i) = W_v x_i$$

After the output o is obtained, a scalar value γ is multiplied to o and residual input x is added to reinforce long-range features that might have lost during the normal convolution layer.

The different in GAN's weakness and SAGAN can be clearly seen in the following figure 3.14.

Figure 3.14: Dog images generated by GAN and SAGAN, demonstrating a weakness in GAN model regarding the integrity of image structure.

3.5 Super-Resolution

3.5.1 Super-Resolution Generative Adversarial Networks

Super-resolution GAN applies a deep network in combination with an adversary network to produce higher resolution images. During the training, A high-resolution image (HR) is downsampled to a low-resolution image (LR). A GAN generator upsamples LR images to super-resolution images (SR). We use a discriminator to distinguish the HR images and backpropagate the GAN loss to train the discriminator and the generator. Figure 3.15 shows the general idea of SRGAN.

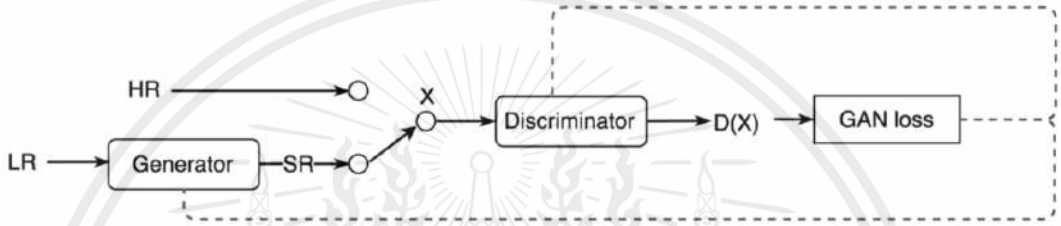

Figure 3.15: A summary of how SRGAN works [26].

Previous super resolution methods are mainly driven by the choice of the optimization function. Most commonly used optimization target for the supervised SR algorithms is the minimization of the mean-squared error (MSE) and maximization of peak signal-to-noise(PSNR) which are defined based on the pixel-wise image differences. The PSNR ratios obtained from these methods is high but the images are perceptually not satisfying.

3.5.2 Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN)

We have seen in the previous subsection how the new method of SRGAN, to use Generative Adversarial Networks for the task of super-resolution is proven to be highly effective. However, there are still some flaws in SRGAN. It is explained in the ESRGAN research paper [30] that the super-resolution images that are generated using SRGAN still possess some unpleasant artifacts such as noise, or grid-like structures. To further enhance the visual quality, the three key components of SRGAN were studied, namely, network architecture, adversarial loss, and perceptual loss, and ESRGAN aims to improve all three of them to derive Enhanced SRGAN (ESRGAN). The ESRGAN paper

[30] introduces the Residual-in-Residual Dense Block (RRDB) without batch normalization as the basic network building unit. The perceptual loss was improved by using the features before activation which means stronger supervision for brightness consistency and texture recovery. Figure 3.16 shows the comparison between the image being generated with SRGAN or ESRGAN and the details can be clearly seen that ESRGAN is better at maintaining the details.

Figure 3.16: A comparison between SRGAN and ESRGAN super-resolution [30].

Wang et al.[30] propose the use of a network where most computation is done on the low resolution image feature space. In order to further improve the recovered image quality of SRGAN, two modifications are made to the structure of the generator G: 1) all BN (Batch Normalization) layers are removed; 2) all the original basic block are replaced with the proposed Residual-in-Residual Dense Block (RRDB), which combines multi-level residual network and dense connections as seen in Figure 3.17

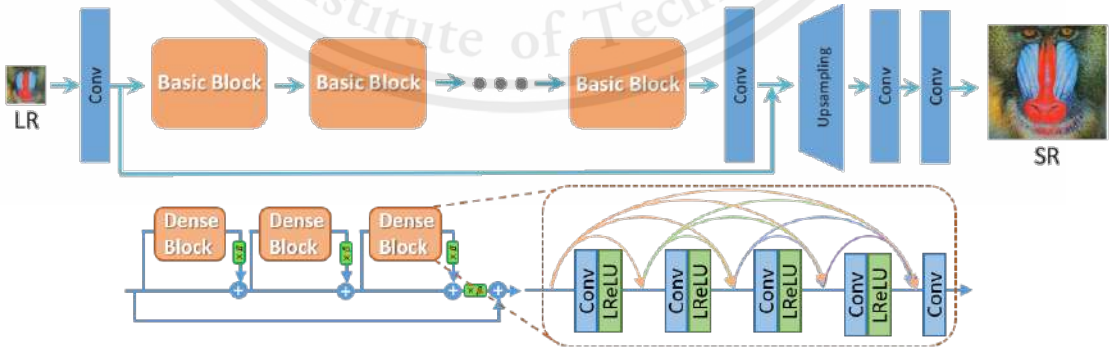

Figure 3.17: The architecture of the ESRGAN (model image from ESRGAN paper)[30].

Chapter 4
Proposed Methods

This chapter comprehensively describes our proposed method, JUAN (Joint Upscaling Adversarial Network) in full detail. Section 4.1 gives an overview of the proposed methods, Section 4.2 explains our attempt on the first method we propose, Section 4.3 elaborates on our new proposed method JUAN, and section 4.4 provides a summary for this chapter.

4.1 Overview

In this section, a new and improved method from Fussell et al. [6] for generating images of galaxies by using a different GAN and Super Resolution model is proposed. The architecture in the previous work uses DCGAN [23] as the generative model and uses another generative adversarial network called StackGAN [24] for the task of super-resolution to increase the resolution of an image from 128×128 to 256×256. We have tried two different methods both using the generative model as SAGAN [8].

4.2 Modified SAGAN with ESRGAN

Our proposed method includes utilizing Self-Attention Generative Adversarial Networks as our base model. SAGAN by Zhang et al. [8] is a special GAN that features a self-attention module at the convolution layer in the network's generative model and discriminative model [8]. Due to the nature of SAGAN's self-attention module, the long-range feature in an image can be constructed that allows the image to retain its structural integrity which is an essential part going forward in galaxy image generation.

We propose that in order to achieve an even higher image fidelity, a super-resolution network, ESRGAN by Wang et al. [30] could be used to upscale the image generated from the SAGAN. Additionally, ESRGAN will be trained with a dataset that

features only the galaxy image in order to reinforce both the generator and discriminator to fully capture the features that are most common amongst galaxies.

4.3 Joint Upscaling Adversarial Network

4.3.1 Modified self-attention mechanism

After experimenting with the base model SAGAN implemented by Park et al.[1] and super resolution for image upscaling, we achieved a modified version of SAGAN in which the following mechanism has been changed.

from:

$$\beta_{ij} = \frac{exp(s_{ij})}{\sum_{i=1}^{N} exp(s_{ij})}, \text{where } S_{ij} = f(x_i)^T g(x_j)$$

to:

$$\beta_{ij} = \frac{exp(s_{ij})}{\sum_{i=1}^{N} exp(s_{ij})}, \text{where } S_{ij} = f(x_i)^T \gamma_f * g(x_j)\gamma_g$$

and from:

$$O_j = v\left(\sum_{i=1}^{N} \beta_{i,j} h(x_i)\right), h(x_i) = W_h x_i, v(x_i) = W_v x_i$$

to:

$$O_j = v\left(\sum_{i=1}^{N} \beta_{i,j} h(x_i)\gamma_h + h(x_i)\right), h(x_i) = W_h x_i, v(x_i) = W_v x_i$$

After experiencing instability of self-attention module, in an attempt to upscale the model to generate 128×128 we propose the inclusion of three extra learnable scalar parameters of γ_h,γ_f, and γ_g introduces a variate rate at which key, query, and value feature map will impact a total change that self-attention module provides in both generative model and discriminative model, which when trained by solely galaxy images should help improves image's clarity and structure integrity of the images.

4.3.2 New discriminator network

During initial training of the modified SAGAN based on the implementation of Park et al. we decide that to change the network of the discriminator of the model. In original discriminator architecture features only convolutional layers as shown below.

[1] https://github.com/heykeetae/Self-Attention-GAN

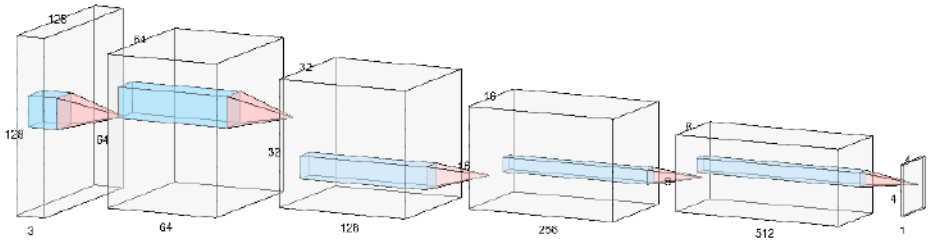

Figure 4.1: Brief overview of a discriminator model inside SAGAN

In our new iteration, we add a global average pulling by Lin et al. [29] layer after convolutional layers in order to flatten our output into a vector that is followed by a fully-connected layer with a scalar output at the end of the network. However, our loss function is now a binary cross-entropy function instead of log loss due to a new sigmoid activator at the final output layer of the fully connected network.

In accommodation with new spatial resolution of 128×128 pixel extra convolutional layers are added comparing to the base SAGAN architecture

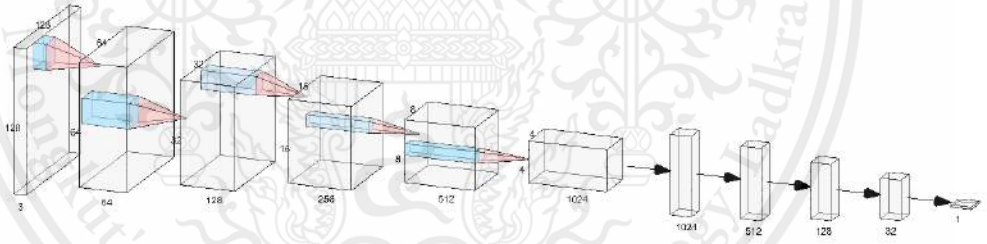

Figure 4.2: Brief overview of a discriminator model inside SAGAN after modification

4.3.3 Super Resolution

For JUAN, since SRGAN [26] is shown to reduce the FID score [25] and the general image quality, we think this is due to not enough SRGAN training . We now use a PSNR driven super resolution method for now, which is peak signal-to-noise ratio based super resolution.

4.4 Proposed Model's Comparison

In this section we will highlight the main different our proposed model, from vanilla SAGAN model [8] and DCGAN + StackGAN model [6]. A full detail of the improvement in the following subsections will be shown and documented later in Chapter 6.

4.4.1 Modified SAGAN vs SAGAN

Modification in self-attention module

As explained briefly earlier in subsection 4.3.1, our SAGAN model inside JUAN adds extra training parameters inspired by the scale parameters in SAGAN [8]. These new parameters inside a self-attention module are trainable scalar variables which will be set to 0 by default, similar to the default SAGAN scalar multiplication. This modification will first let the self-attention module solely rely on neighboring data first, and then gradually each latent feature map of key, query, and value will be more prominent due to the change in their respective scalar multiplication. This allows key and query to have a different degree of prevalence in the attention map, which affects the overall effectiveness of the self-attention module. We hypothesize that the different degrees to which key, query, and value are related might be a key toward generating better galaxy images with finer detail and clearer structure compares to normal SAGAN. See the following figure 4.3 for a brief overview of the modified model.

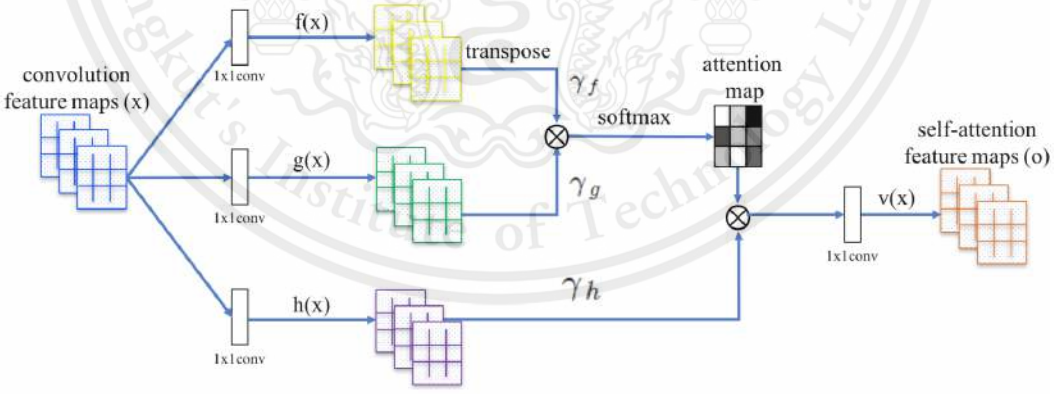

Figure 4.3: Brief overview of a modified SAGAN module

4.4.2 JUAN vs DCGAN + StackGAN

Proposed by Fussell et al.(2019) [6], using DCGAN for galaxy image generation and upscaling the generated image by StackGAN. One weakness we can see with DCGAN + StackGAN model is that DCGAN may not be able to keep the NN distinct features such as spiral arms when generating an image. Our proposed method of using a modified SAGAN and a super-resolution network does preserve these distinct features such as spiral arms, as explained in Section 3.3.6 how SAGAN is the best choice for generating these features. We show the difference between our proposed method and the method proposed by Fussell et al. [6] in chapter 6 Section 6.1.

4.4.3 ESRGAN for super resolution

ESRGAN has recently been tested and used with our model, it has performed significantly better than all of our previous methods such as PSNR and SRGAN, and the main problem was the SRGAN wouldn't train properly which is probably because of the low quantity of training data that we provide to the SRGAN. Without training ESRGAN from scratch, we just use a pretrained ESRGAN and train the galaxy training dataset over that which produced very great results, the resolution is higher and the grid from the SAGAN generation is lost which means the image has much less noise and is a lot clearer, so we will try to modify ESRGAN by adding a method to compensate the low training data such as augmented discriminator and combine the ESRGAN + a 128 version of SAGAN

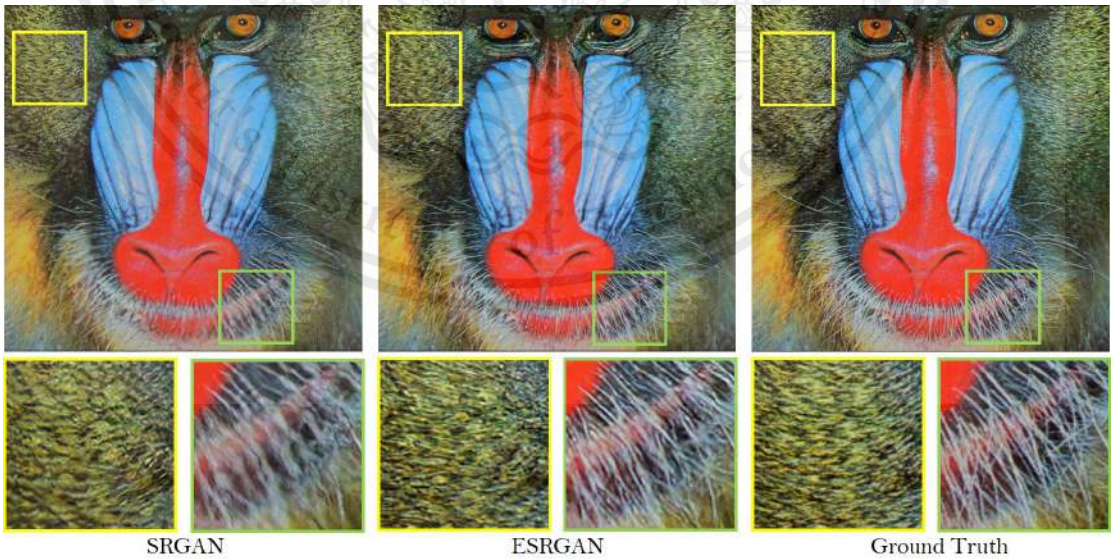

Figure 4.4: ESRGAN comparison with normal super resolution GAN and ground truth

4.5 Summary

To summarise the current working method that we have proposed is the modified version of SAGAN and a super resolution module which uses ESRGAN for upscaling and combing the SAGAN and super resolution we have a joint upscaling adversarial network (JUAN). Figure 4.5 shows the steps of our proposed method.

- 1.A resolution 128×128 image is generated using the SAGAN.
- 2. A super resolution module (ESRGAN in this case) which upscales and enchances the details of the generated low resolution image to a 256×256 image.
- 3. A post-prosessing algorithm that adjusts the brightness, saturation, contrast, sharpness and other parameters to enhance the image and reduce any noise from the GANs in the previous step.

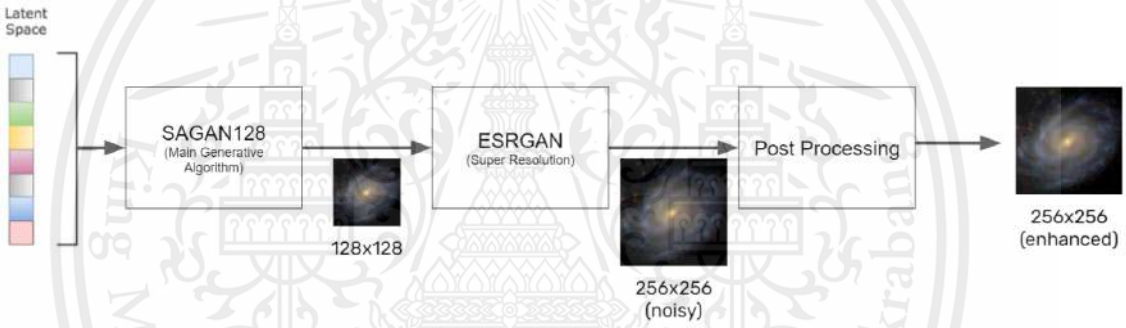

Figure 4.5: A graphical representation of our proposed method.

Chapter 5
Implementation

This chapter focuses on the methodology and implementation of our proposed methods including the generative part, and the super-resolution part. We also explain the dataset we used in the process of training and their sources.

5.1 Dataset

5.1.1 Galaxy Zoo

The datasets we use to train the SAGAN model are primarily from Galaxy Zoo [20]. The images from Galaxy Zoo are various galaxies with different classifications and categories. Galaxy Zoo classifies its vast collection of galaxy images via a volunteering process from humans by letting their users select which type of galaxy they are looking at on Galaxy Zoo's web service interface. All of the images then are stored within Galaxy Zoo's database which can be accessed. However, Galaxy Zoo images are primarily used for classification-related tasks rather than a generative task, we opt to prepare our own dataset comprised of smaller quantities but higher clarity and resolution called the GalaxyMan dataset.

5.1.2 GalaxyMan

GalaxyMan was made in an attempt to solve the problems with the GalaxyZoo dataset, such as the lack of clarity in the distinct features of a galaxy in the GalaxyZoo images such as the clarity of spiral arms of a galaxy GalaxyMan dataset focuses on the image of spiral galaxies in particular to determine SAGAN's capability of generating long-range features such as the structure of the spiral arm. We handpicked images from various sources such as space.com[1], eso.org[2], astronomy.com[3] and most importantly

[1] https://www.space.com/
[2] https://www.eso.org/public/
[3] https://astronomy.com/

apod.nasa.gov[4]. The original GalaxyMan only had 66 images but was then increased by handpicking some GalaxyZoo images that were usable with clear features, 15% of the images are from GalaxyZoo which in the total count to about 104 images total for GalaxyMan. Two following figures 5.1 and 5.2 display a sample of 5 images from GalaxyMan in stark with 5 images from GalaxyZoo respectively.

Figure 5.1: Sample images from GalaxyMan dataset

Figure 5.2: Sample images from GalaxyZoo dataset

5.1.3 SpiralZoo

For the experiment's purpose, we also use the third dataset, SpiralZoo is created by combining images from 2 prior datasets 5.1.1 and 5.1.2. SpiralZoo is comprised of 96 images with each half from GalaxyMan and GalaxyZoo. This combination of the two shows the best results of the experiment with the dataset can be seen in chapter 6, table 6.5.

The following table 5.1 captures the quantity of images used in the training of the model from each dataset.

Dataset	GalaxyZoo	GalaxyMan	SpiralZoo
Images	79977	104	96

Table 5.1: SpiralZoo dataset is a dataset of mixed images from both Galaxy Zoo and GalaxyMan. Images selected from Galaxy Zoo in SpiralZoo dataset are more feature prominent than other Galaxy Zoo images

[4]https://apod.nasa.gov/apod/astropix.html

5.2 Data Pre-processing

The implementation is done by using a python machine learning and computer vision framework PyTorch. In PyTorch, an image will be stored in a form of a tensor, a three-dimensional matrix that stores elements of the same data type. Images will be resized to the desired spatial dimension first before being used in the models.

```
tensor([[ 3.1416,  3.1416,  3.1416,  3.1416],
        [ 3.1416,  3.1416,  3.1416,  3.1416],
        [ 3.1416,  3.1416,  3.1416,  3.1416]], dtype=torch.float64)
```

Figure 5.3: A float-64 tensor object implementation by PyTorch

5.3 SAGAN

A pretrained SAGAN implementation by Park et al.[5] is used for training the galaxy generative model. The preferred image spatial size in this particular implementation is 64×64 pixel per image. Unfortunately, our attempt to readjust the image size that can be generated by this model to 128×128 pixel encountered a mode collapsing problem during the training session where a supposedly low-quality image generated by the generative model can outperform a discriminator which leads to a halt in model parameter learning resulting in unrecognizable images. This problem is circumvented by the usage of a super-resolution model on the generated 64×64 images.

However, we have achieved a better resulted on the generated galaxy images on Frechet Inception Distance score [25] compared to the base model of SAGAN at 64×64 resolution which will be explained in detail in the next chapter by using our proposed version of SAGAN with a training batch size of 16. In our modified version of the model, Joint Upscaling Adversarial Network, three extra learnable scalar parameters are multiplied to the key, query, and value latent feature maps in the self-attention module.

We have converted some SAGAN modules to classes and our modified SAGAN is implemented by this class. Fig 5.4 shows the UML class diagram of our modified SAGAN.

[5] https://github.com/heykeetae/Self-Attention-GAN

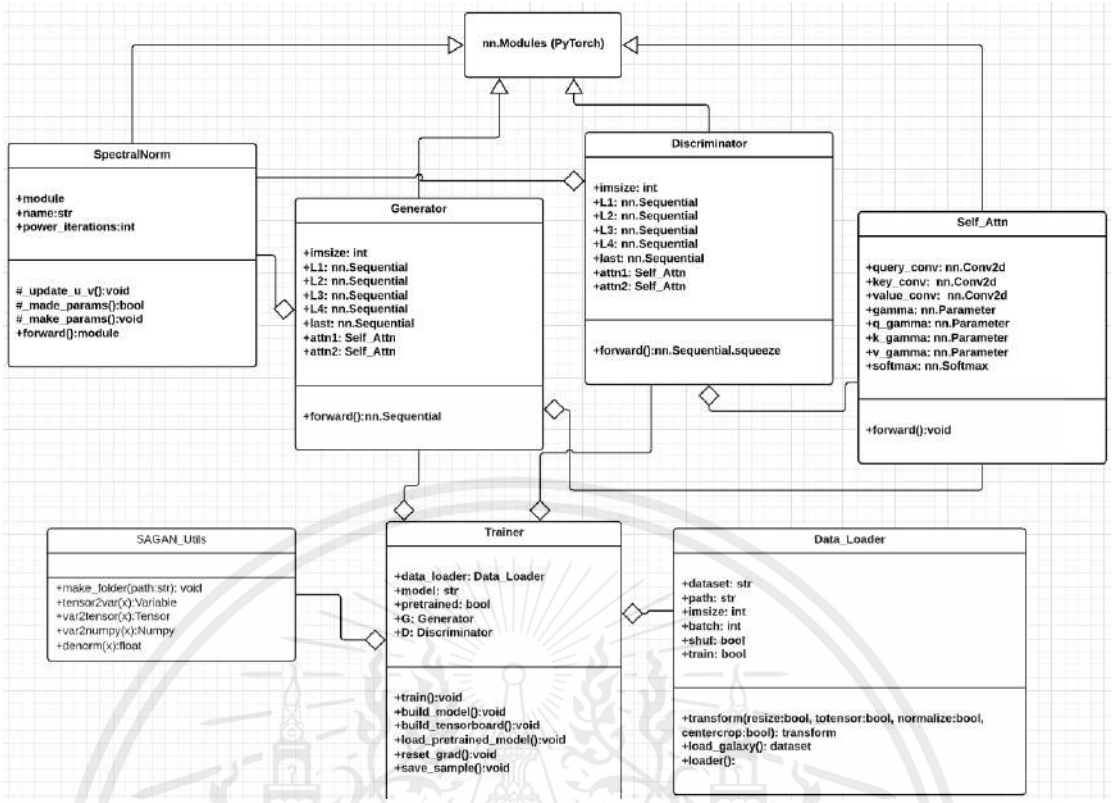

Figure 5.4: Class diagram of our modified SAGAN

5.4 Super Resolution

An implementation of the super resolution model, both for PSNR based super resolution and SRGAN [6] is used for training the SRGAN and PSNR based super resolution

5.4.1 Training of SRGAN

The training of SRGAN, required four directories of the dataset

- High resolution images training directory,
- High resolution images validation directory
- Low resolution images validation directory
- Low resolution images validation directory

[6] https://github.com/idealo/image-super-resolution

The training time for the SRGAN was about 2 days, but unfortunately, the testing of the trained model did not show promising results, so we had to halt using SRGAN temporarily.

5.4.2 ESRGAN

As explained in chapter 3, ESRGAN [30] is superior in visual quality to SRGAN and all other methods we have applied and tested. ESRGAN (Enhanced Super Resolution GAN) has been shown to decrease noise and upscale the image without losing main features and qualities.

The training for ESRGAN took approximately 2 days in total because we are continuing to train from a pretrained model provided by the official ESRGAN repository. ESRGAN is able to upscale the image up to 4 times.

5.5 Graphical User Interface

This section shows the simple graphical user interface implemented for the ease of use in generating the images in batches. In the scope of this research currently, We did not consider to implement a beautiful user interface or focus on the design or the modern architecture of user interfaces, we have instead focused on the ease of use in which the generation will not be done by typing and running the python model in a command line interface.

5.5.1 GUI implementation

We have used pyqt5 library which is also compatible with Python for the creation of the GUI.

The GUI has been designed using the pyqt5 designer which outputs a (*.ui) file which can be converted to a python class file using the inbuilt pyqt commands

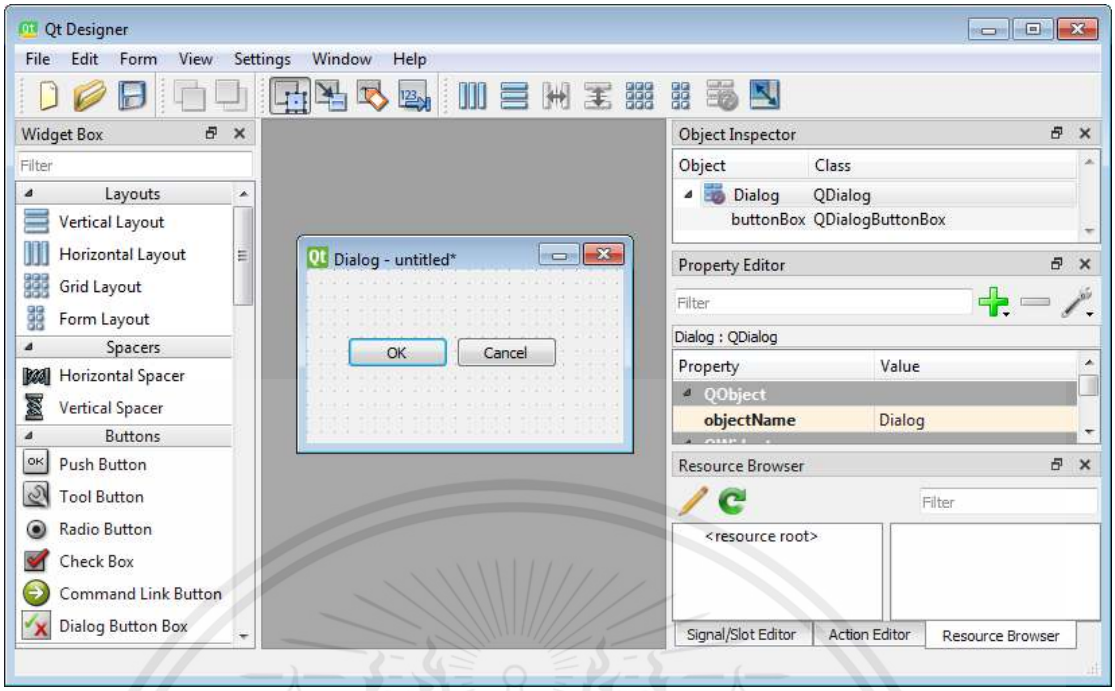

Figure 5.5: An example of the qt designer interface.

Fig 5.5 shows the basic user interface of the qt5 designer form, using this qt design a simple GUI was created but the functionalities and the on click method must be coded after the GUI form is generated.

5.5.2 Final GUI

The final GUI that was created is a very simple GUI only meant to be used for a simpler user experience of generating the galaxy images instead of having to input each parameters through command line.

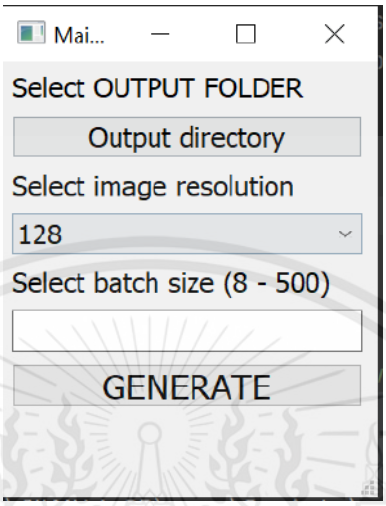

Figure 5.6: The final GUI.

Fig 5.6 shows the final GUI of our project, and was developed and implemented only for the ease of use and to avoid running the generation from command line. The user must enter the file location they want the output galaxy images to be saved and the batch size of how many image will be generated(quantity) and the resolution of each generated image (128, 256 or 512).

5.6 Testing and verification

5.6.1 Unit test: design and implementation

The first type of test we are choosing to execute is the unit tests for some of the supporting functions of the main program, such as functions that convert matrices or tensors to different data types such as Numpy array or Python Lists.

We chose to test the class Utils, as it is a class which is used by every other class or method, as it does the basic conversion of tensors and basic matrix operations, it is absolutely essential to test .

Funnction definition	Function description	Output
tensor2var(x:torch.Tensor)	Converts Pytorch Tensor to Pytorch Variable	torch.Variable
var2tensor(x: torch.Variable)	Converts Pytorch Variable to Pytorch Tensor	torch.Tensor
var2numpy(x: torch.Variable	Converts to a numpy compatible array	numpyArray
denorm(x: torch.Tensor)	Used to denormalize a Tensor	torch.Tensor
generate_zeros(n: int)	Generates a Tensor filled with 0 with length n	torch.Tensor
generate_ones(n: int)	Generates a Tensor filled with 1 with length n	torch.Tensor

Table 5.2: The functions of the Utils class.

The functions of the Utils class files are as shown in the table **??**:

5.6.2 Tools: Pytest vs Unittest

Firstly we would like to start by reporting what we have learned about some tools during this project. Our project is almost entirely coded using the Python programming language, therefore we will be discussing some of the tools we used during the tests and the advantages or disadvantages of each of them.

Unittest is the library provided by Python for simple and quick implementation of Unittest. Pytest on the other hand is a third party library which is open source and can be installed using standard pip command.

One of the most obvious advantages of Pytest is the fixtures decoration over the classes methods or standalone functions of test cases that can provide a much better flexibility and relaying and displaying information over just using print statements in the respective code block.

Unittest is easier to read and much more friendly for new testers and it works out of the box, but once we get familiar with Unittest the other advantage of Pytest is the fact that Pytest can test on the classes that were created for Unittest, Pytest is compatible with Unittest.

For the intents and purposes of the test activities conducted this round, we have decided to choose Pytest to familiarize ourselves with a more robust and more industrially accepted testing library for python.

Chapter 6
Experimental Result

In this chapter, we will show the generated images from our model and the images generated from the related work (Fussell et al.(2019) [6]) in Section 6.1 and give a visual comparison before we move on to elaborate on the experiments we have conducted to evaluate our proposed method. We have conducted two different experiments to evaluate our model. The first experiment is the FID (Fréchet Inception Distance) score, which is elaborated in more detail in Section 6.2, and the second experiment is the Visual Turing Test. Both of these experiments will be elaborated in each Phase, Phase I represents the progress and the experiments we conducted in the first semester, and Phase II is the progress after the second semester> Each Phase is presented in Sections 6.3 and 6.4 respectively. In Section 6.5 we will elaborate on the importance of using the SpiralZoo dataset and the experiments we conducted relating to the dataset and why it showed the best results while training.

6.1 Visual comparison

6.1.1 Galaxy images generated from DCGAN+StackGAN Fussell et al. [6]

Here we have randomly chosen 5 images to display the generated images from the proposed method of Fussell et al. [6] which is DCGAN for image generation and StackGAN for upscaling of the image as shown in fig 6.1.

Figure 6.1: Example generated image from DCGAN with StackGAN

6.1.2 Galaxy images generated from our proposed method (JUAN)

Here we have randomly chosen 5 images to display the generated images from our proposed method, which is a modified SAGAN for image generation and PSNR for applying super-resolution to the generated image as shown in fig 6.2.

Figure 6.2: Example generated images from our new proposed method (JUAN)

As seen from fig 6.2 features of spiral galaxies can be seen such as the spiral arms from the images of galaxies that are generated from our proposed method, these features are now more visible and more high resolution, even more than our previous proposed method using the SAGAN64 + PSNR which is now (SAGAN128 + ESRGAN).

6.2 Fréchet Inception Distance

FID [25] is a measure of similarity between two datasets of images. It was shown to correlate well with the human judgment of visual quality and is most often used to evaluate the quality of samples of Generative Adversarial Networks. FID is calculated by computing the Fréchet distance between two Gaussians fitted to feature representations of the Inception network. To simply understand the FID score, one could think of it as the following: a lower FID score means less difference between the original dataset and the generated dataset, hence the lower the FID score the better.

6.3 Phase I

6.3.1 FID Experiment Setup

Since FID works by comparing two datasets, we had to prepare a dataset that is comprised of images generated from the generative model in that experiment exclusively. So one input would be the dataset we used to train the generator which for this experiment SpiralZoo256 was used, and the other input will be the same number of images as the dataset but the images are generated from the generative model (SAGAN or JUAN).

We divide the FID experiment into two parts, the first part is by using SAGAN and the second part is by using JUAN. There are 8 experiments in total:

- Default SAGAN: using the default SAGAN with no modification.
- SAGAN+SRGAN: using SAGAN along with the trained SRGAN to upscale the image generated from SAGAN.
- SAGAN+PSNR: using SAGAN and PSNR based super resolution to upscale the image generated from SAGAN.
- JUAN with no super resolution: Using the modified SAGAN without super resolution.
- JUAN+PSNR: using JUAN with PSNR based super resolution.
- JUAN+PSNR+denoise=1 : using JUAN with PSNR based super resolution and using a fast local means a denoising factor of 1 before inputting the generated image to the super resolution.
- JUAN+PSNR+denoise=2 : using JUAN with PSNR based super resolution and using a fast local means denoising factor of 2 before inputting the generated image to the super resolution.
- JUAN+PSNR+denoise=3 : using JUAN with PSNR based super resolution and using a fast local means denoising factor of 3 before inputting the generated image to the super resolution.

Method id	GAN	Super Resolution	denoise
1	SAGAN	-	-
2	SAGAN	SRGAN	-
3	SAGAN	PSNR	-
4	JUAN	-	-
5	JUAN	PSNR	-
6	JUAN	PSNR	1
7	JUAN	PSNR	2
8	JUAN	PSNR	3

Table 6.1: Summary of the 8 tests we conducted in the FID experiment

Table 6.1 summarizes the tests we conducted in our experiment to calculate the FID score for each method.

6.3.2 FID Experimental results

Here we will show the experimental results of the 8 experiments we have set up.

Table 6.2 shows the FID score calculated for each of the methods described above. Method 7 (JUAN + PSNR + denoise=2) has the best FID score which is the lowest score with an FID score of 218.

We can clearly see in table 6.2 when SRGAN [26] is used for super-resolution the FID goes up which indicates a bad performance and that the image generated is not similar to the dataset, we suspect this is because not enough training of the SRGAN and this is why we are temporarily using PSNR based super-resolution instead of SRGAN.

Method id	GAN	Super Resolution	denoise	FID
1	SAGAN	-	-	245
2	SAGAN	SRGAN	-	294
3	SAGAN	PSNR	-	232
4	JUAN	-	-	225.67
5	JUAN	PSNR	-	225.6
6	JUAN	PSNR	1	225
7	JUAN	PSNR	2	218
8	JUAN	PSNR	3	230

Table 6.2: Summary of the 8 tests we conducted in the FID experiment and the FID score corresponding to each test

Figure 6.3 shows a graphical representation of the FID corresponding to each experiment. NOTE: Y AXIS STARTS FROM 200 for easier comparison between the models.

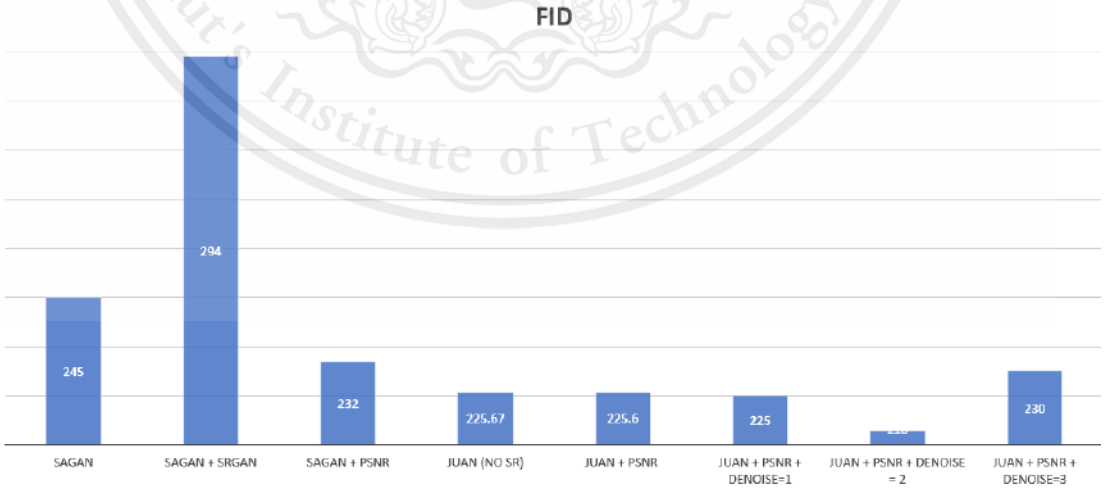

Figure 6.3: A graphical representation of the FID corresponding with each experiment

6.3.3 VTT Setup

We have created a survey with Google Forms[1] where there are 5 sections in the survey and 7 images in each section.

There are at most two real images of galaxies and the rest are generated images of galaxies (fake images) which were generated from a generative model. In the five sections, four sections are the methods for which we tested the FID, with the best results in each method, table 6.3 shows the method we have used in each section and the corresponding FID to those methods. The five sections are divided as the following:

- The fake images are generated from the default SAGAN (FID: 245).
- The fake images are generated from the default SAGAN + PSNR based super resolution (FID: 232).
- The fake images are generated from the JUAN without super resolution(FID: 225).
- The fake images are generated from the JUAN + PSNR based super resolution (FID: 218).
- The fake images are generated from the model of the previous paper, Forging New worlds which used DCGAN + StackGAN for super resolution [6].

Method id	GAN	Super Resolution	denoise	FID
1	SAGAN	-	-	245
3	SAGAN	PSNR	-	232
4	JUAN	-	-	225.67
7	JUAN	PSNR	2	218
-	DCGAN	StackGAN	-	-

Table 6.3: Summary of the 5 Sections of the visual Turing test and the FID score corresponding to each test

6.3.4 VTT Experimental results

The survey has been posted to many astronomy and astrophysics-related forums on sites like Reddit [2] or Facebook groups. We have stopped accepting responses to this

[1] https://www.google.com/forms/about/
[2] www.reddit.com

survey since 17th December. There are in total 104 responses. Table 6.4 and Fig 6.4 shows the percentage of people who thought the fake image is real for each method of image generation and the corresponding FID score for each method.

Method	SAGAN	SAGAN+PSNR	JUAN	JUAN+PSNR	DCGAN+StackGAN
Percent	53.59	46.28	46.60	55.83	38.67
FID	245	232	225.67	218	-

Table 6.4: Percentage of people who thought the fake image is real for each method of image generation

Since the lower, the FID score is the better the GAN is, and JUAN+PSNR has the lowest FID score and the highest percent of people who thought the fake image was real as seen from table 6.4. On the contrary the method from Fussell et al. [6] which uses DCGAN + StackGAN [6] had only around 38% of people who thought the fake image was real.

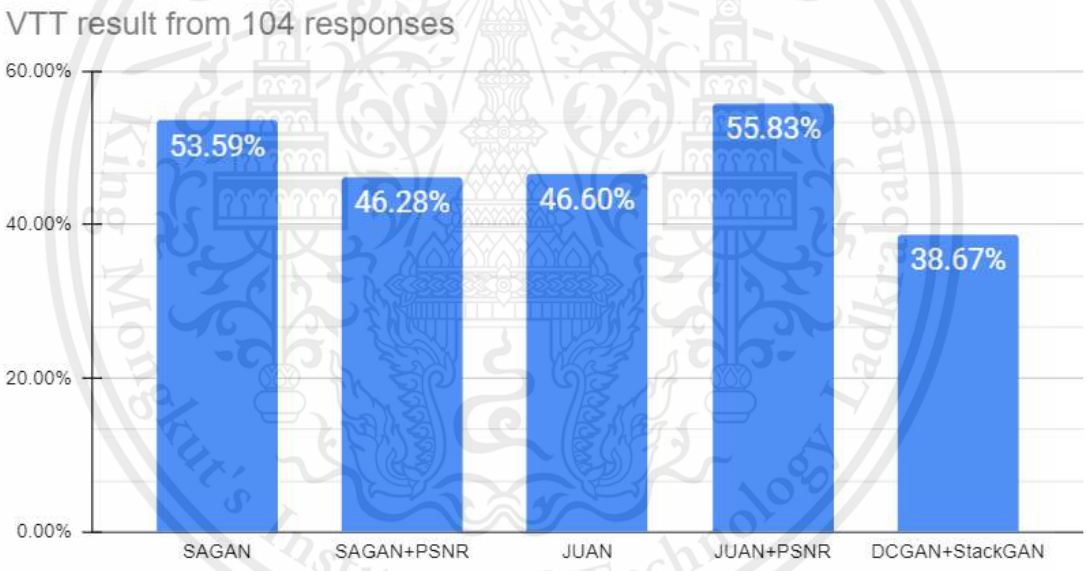

Figure 6.4: A graphical representation of the percentage of people who thought the fake image is real for each method of image generation

6.4 Phase II

6.4.1 FID Experiment Setup

The setup for the FID experiment in Phase II is similar to the setup in Phase I but with one main difference, the test cases of the generated datasets were not selected at random, but selected using the pairwise combinations of the different parameters that can use to generate the galaxy images.

The optimization we are working on is the optimization of the parameters of the super-resolution and the image processing algorithms, these parameters will all affect the output image in a different way, and we need to find a combination that will minimize the FID score which means the best image output. We have the following controllable parameters:

- brightness of the image,
- contrast of the image,
- Sharpness of the image,
- OpenCV Denoise function factor,
- Gaussian blur,
- Color (HSV),
- Image resolution,

The combination of these different parameters may produce better results than other combinations, and the resulting image can be testing using the FID (The Fréchet inception distance (FID) is a metric used to assess the quality of images created by the generator of a generative adversarial network and closely resembles human judgment), the lower the FID the better, we will aim to find the min FID by using different combinations of the parameters mentioned above.

The parameters and the partition of each value are as follow:

In this case, the higher resolution will usually cause the resulting image to have more noise than the lower resolution ones. Some values of each parameter were not considered to be tested at all as those values may completely remove some important features from the image, such as a high blur or denoising factor or a high color saturation factor.

KEY	Description	Partition
b	Brightness factor	medium=1.2 high=1.5
c	Contrast factor	medium=1.5 high=2
s	Sharpness factor	medium=1.5 high=2
d	Denoise factor	medium=2
gb	Gaussian blur factor	low=7 medium=9 high=11
col	Color saturation factor	medium=1.2
resolution	Resolution of the image	low=128 medium=256 high=512

From the above we can see that we have 7 parameters in total and each parameter may have 2 or 3 values to test, which means the total combination would be: 2x2x2x1x3x1x3 = 72 different combinations, which is a lot to test

Therefore we have decided to go with a pairwise combination of these parameters which results in 22 different combinations that we need to test. The 22 different pairwise combinations are as follows:

Test ID	denoise	gblur	brightness	color	constant	sharpness	resolution
1	2	7	1.5	1.2	2	1.5	512
2	2	11	1.5	1.2	2	2	128
3	2	7	1.2	1.2	1.5	1.5	256
4	2	9	1.5	1.2	2	2	512
5	2	7	1.2	1.2	1.5	2	128
6	2	9	1.5	1.2	2	1.5	256
7	2	11	1.2	1.2	1.5	2	512
8	2	7	1.5	1.2	2	1.5	128
9	2	9	1.2	1.2	1.5	2	256
10	2	11	1.5	1.2	2	1.5	512
11	2	9	1.2	1.2	1.5	1.5	128
12	2	11	1.5	1.2	2	2	256
13	2	7	1.2	1.2	1.5	1.5	512
14	2	9	1.5	1.2	2	2	128
15	2	11	1.2	1.2	1.5	1.5	256
16	2	7	1.5	1.2	2	2	512
17	2	11	1.2	1.2	1.5	2	128
18	2	7	1.5	1.2	2	1.5	256
19	2	9	1.2	1.2	1.5	2	512
20	2	11	1.5	1.2	2	1.5	128
21	2	7	1.2	1.2	1.5	2	256
22	2	9	1.5	1.2	2	1.5	512

6.4.2 FID Experimental results

The results of this pairwise testing were very insightful as it provided some combination of parameters with the best FID output, the output shown below corresponds to the test ID shown in the table above. The minimum FID output was from test case 3, as shown in the table below.

Test ID	1	2	3	4	5	6	7	8	9	10	11
Test FID result	239	250	219	229	260	231	225	262	253	229	258

Test ID	12	13	14	15	16	17	18	19	20	21	22
Test FID result	245	236	253	255	223	251	230	243	257	238	239

6.4.3 Visual Galaxy Test v2

We have decided to overcome the concerns of the previous experiments such as the concern that the participants of the survey may not have the essential knowledge of astronomy or astrophysics to contribute to the survey and might give false or biased results. We have designed a new Visual Galaxy Test which will be slightly different, instead of dividing the different images into different sections, we will randomly place the images and let the participants choose ”Real” or ”Fake” based on their knowledgeable opinion of if the galaxy images look artificial or not. Before the main test, however, each participant must go through a series of small quizzes that will determine their level of understanding of astronomy or astrophysics, and we can filter out the results of participants with a score in the quiz which is less than 80%.

6.4.4 Visual Galaxy Test: Astronomy quiz

The quiz that will test the participant's general knowledge in the field of astronomy may help us filter out the biased and false results from participants who might not have the adequate knowledge to identify a real galaxy properly, which in turn may label the artificial images with using intuition instead of knowledge.

All questions will be multiple choice except the first question which asks the participant about their knowledge level. The questions designed initially were text based questions such as ”What is a galaxy”, and we realized that such questions could be easily searched and the answers might not be their true knowledge. The questions of the quiz will be as follows:

- 1. Please rate your knowledge of astronomy on a scale of 1-10, where 1 is the least knowledge and 10 is higher education level understanding of astronomy.

- 2. What is the following structure?

Figure 6.5: Image from question 2 of the quiz.

  - A. A bright star like object that is near the solar system.
  - B. A huge collection of gas, dust, and billions of stars and star systems all bound by gravity.
  - C. A type of black hole which is brighter than usual.
  - D. The observable universe.

- 3. What type of galaxy is the following?

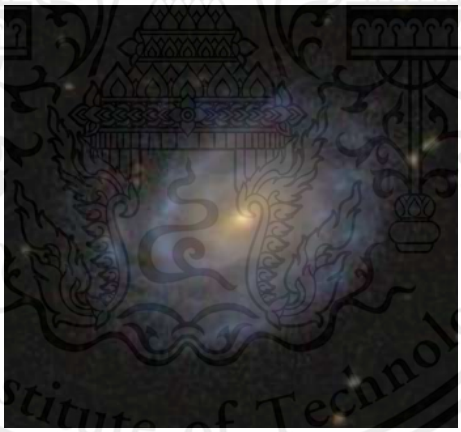

Figure 6.6: Image from question 3 of the quiz.

  - A. Elliptical.
  - B. Spiral.
  - C. Irregular.
  - D. This is not a galaxy.

- 4. What is the state of the nucleus of this galaxy?

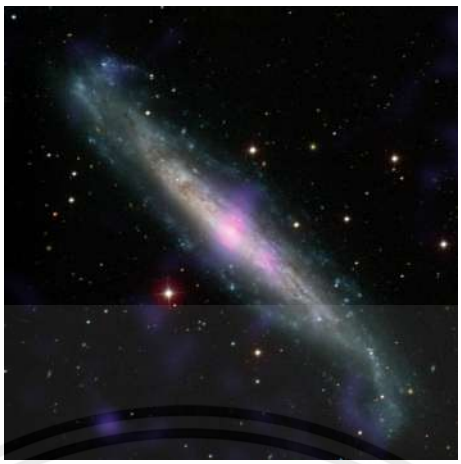

Figure 6.7: Image from question 4 of the quiz.

- A. There exists a massive wormhole in the middle of the galaxy.
- B. Active Galactic Nuclei (AGN).
- C. The density in the middle of this galaxy is almost 0.
- D. This is not a galaxy.

- 5. What type of galaxy is the following?

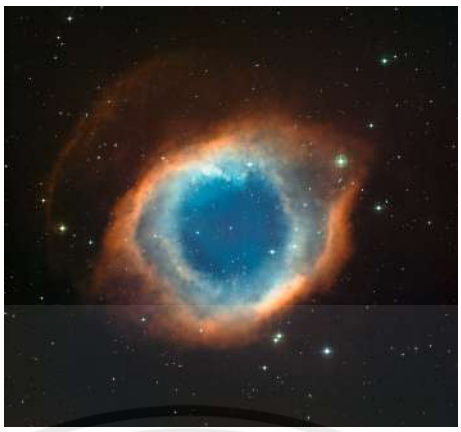

Figure 6.8: Image from question 5 of the quiz.

  - A. Spiral, because of the disk like features.
  - B. Spiral, because of the color difference between the middle and the edge.
  - C. Elliptical, because of the rounded structure.
  - D. This is not a galaxy.

- 6. What type of galaxy is the following?

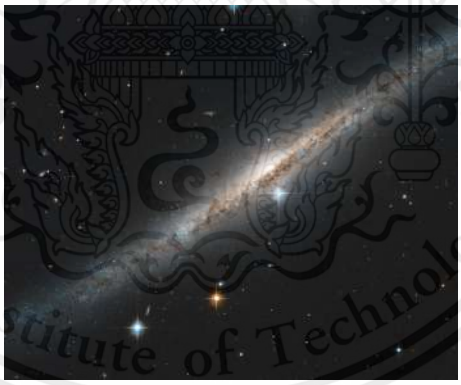

Figure 6.9: Image from question 6 of the quiz.

  - A. Spiral.
  - B. Elliptical.
  - C. This is not a galaxy.

6.4.5 VGT Experimental Setup

This time, we are changing our methods slightly. Last time we showed the images in sections, where each section represents each algorithm, so to remove more bias and get a more normalized result, we have removed the use of sections and just randomized the order of images, we are currently using 4 algorithms to generate 5 images from each of them and 1 real galaxy image, so a total of 20 synthetic galaxy images and one real galaxy image.

The four algorithms we used to generate the images are the following:

Algorithm name	**Description**
SAGAN	The very original SAGAN from the original paper's repository (Goodfellow et al., 2019)
DC + STACKGAN	The related work which uses the DCGAN + STackGAN, (Fussel et al., 2019)
JUAN64	Our proposed method from semester 1 (SAGAN64+PSNR)
JUAN128	Our current and recent proposed method (SAGAN128+ESRGAN+post processing)

6.4.6 VGT Experimental results

Out of the 91 participants, 71 answers were collected and calculated and 20 answers were ignored. The 20 answers were ignored because they failed to pass our criteria of basic understanding in astronomy.

Algorithm name	**Percent of participants who thought the fake image was REAL for each algorithm**
SAGAN	30.3%
DC + STACKGAN	22.2%
JUAN64	43%
JUAN128	**75.2%**

Table 6.5 shows the name of the algorithm on the left and the percentage of the participants who thought that the synthetic galaxy image was an image of the real galaxy on the right. Our recent proposed method got over 75% of participants to believe that the synthetic image is a real galaxy, and these participants are no random participants as shown earlier, they have the knowledge required in astronomy to distinguish between galaxy types, states, and composition.

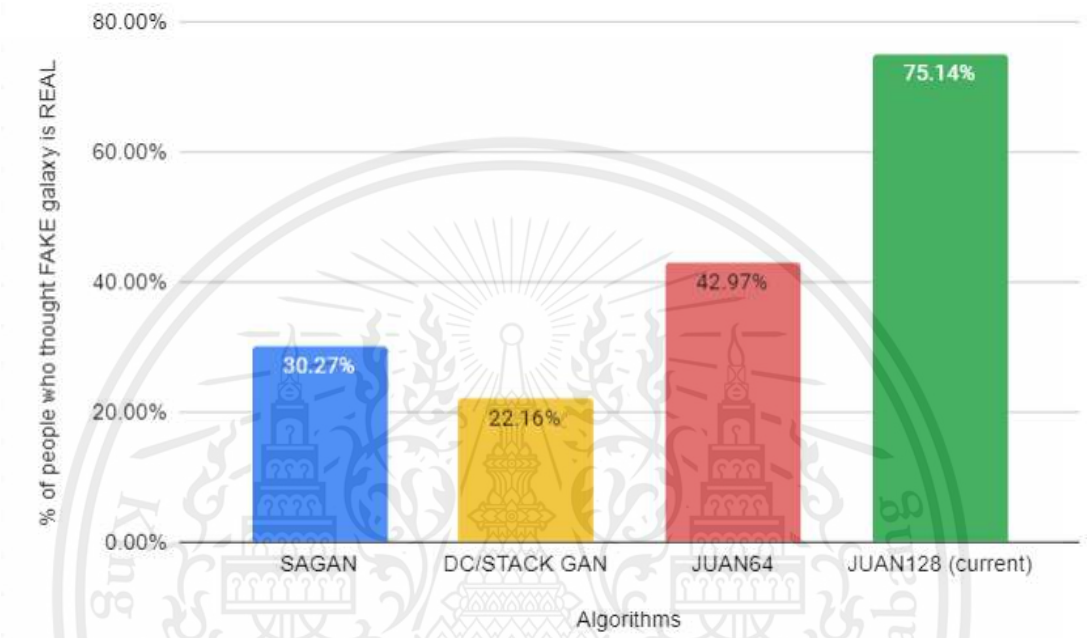

Figure 6.10: Bar graph representing recent VGT results.

Fig 6.10 shows the bar graph representation of the percentage of people who thought the synthetic galaxy image was real in each algorithm, our recent algorithm is on the rightmost side which is the JUAN128.

6.5 Dataset experiments

Images from datasets we used can be seen in chapter 5 from fig 5.1 and 5.2. Here we will show why we chose to use the SpiralZoo dataset which is a combination of both GalaxyZoo and GalaxyMan dataset.

Table 6.6 shows that when SpiralZoo dataset is used to train the same SAGAN model it gives the best FID score with over 100 score difference than using GalaxyMan or GalaxyZoo, and this is the reason we chose SpiralZoo dataset to train our models.

Dataset	GAN	FID
GalaxyMan	SAGAN	345
GalaxyZoo	SAGAN	386
SpiralZoo	SAGAN	245

Table 6.5: Comparison of datasets used to train the GAN model and the FID score corresponding to the different dataset

6.6 Summary

The experiments show that our newest model with three different components, the Main modified version of SAGAN128 + ESRGAN trained on the low and high-resolution galaxy dataset + image processing algorithms or in short our current and the recent method that we propose, JUAN128, has seen vast improvements from the previous semester's proposed method. The image from JUAN128 is of higher resolution and higher quality and has proven to confuse over 75% of participants and making them believe that the synthetic GAN-generated image is a real image.

Chapter 7
Conclusion

7.1 Conclusion

Currently, with the new Joint Upscaling Adversarial Network as our proposed method, we have a result that is better in comparison with the base SAGAN model both with super-resolution enhancing the output images in FID scores [25]. After updating our model to achieve 128×128 resolution, our model's training suffers from an increase in parameters of the larger model with not enough amount of training dataset, resulting in fewer variant images produced by the model, however, still retain better image quality and spatial sizes than before.

The current implementation of super-resolution model within our proposed are also specially trained by upsampling galaxy images to 256×256 along with various computer techniques such as denoise, Gaussian blur certainly help push the overall quality of the final images we can achieve, contributing to an even better FID score when being compared to the images without extra image-enhancing algorithms.

7.2 Recommendation for future work

With our new generative model capable of producing image 128×128, it is merely a first step in generating next-generation galaxy images towards astronomical research contribution. Unfortunately, due to our hardware budget limitation and lack of resources for the significantly larger dataset, We humbly recommend that experimenting with the model's architect, hyperparameters testing, and gather an extremely larger training dataset will go a long way in achieving even better results. We also felt that by adding auxiliary value in both the generative model and super-resolution model to achieve conditional GAN model and being able to generate irregular type also, we found that due to the nature of irregular galaxies being simply a blight circular, singled-toned object, it is hard for the models to pick up details on the overall structural shape of the galaxy.

Lastly, myriads of other image-enhancing algorithms maybe are experimented with generated images from the models that we have not tried yet may be key to improving some of the model's flaws caused by the limited numbers of high-quality galaxy images that can be gathered for the training dataset. Each unexplored combination of image post-processing algorithm with different parameters may yield better image clarity, clearer details of the end result.

Bibliography

[1] , H.Kragh,Expanding Earth and Static Universe: Two Papers of 1935. [PDF] Available at: <https://arxiv.org/abs/1507.08040> [Accessed 17 August 2020]

[2] P. J. E. Peebles,Bharat Ratra, The Cosmological Constant and Dark Energy. [PDF] Available at: <https//arxiv.org/abs/astro-ph/0207347> [Accessed 21 August 2020]

[3] Diederik P Kingma, Max Welling, Auto-Encoding Variational Bayes. [PDF] Available at: <https://arxiv.org/abs/1312.6114> [Accessed 15 August 2020]

[4] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, Generative Adversarial Networks. [PDF] Available at: <https://arxiv.org/abs/1406.2661> [Accessed 15 August 2020]

[5] Siamak Ravanbakhsh, Francois Lanusse, Rachel Mandelbaum, Jeff Schneider, Barnabas Poczo, Enabling Dark Energy Science with Deep Generative Models of Galaxy Image. [PDF] Available at: <https://arxiv.org/abs/1609.05796> [Accessed 14 August 2020]

[6] Levi Fussell, Ben Moews, Forging new worlds: high-resolution synthetic galaxies with chained generative adversarial network. [PDF] Available at <https://arxiv.org/abs/1811.03081> [Accessed 14 August 2020]

[7] Edwin Hubble, Extra-Galactic Nebulae. [PDF] Available at: <http://articles.adsabs.harvard.edu/pdf/1926CMWCI.324....1H> [Accessed 7 september 2020]

[8] Han Zhang, Ian Goodfellow,Dimitris Metaxas, Augustus Odena, Self-Attention Generative Adversarial Networks. [PDF] Available at: <https://arxiv.org/abs/1805.08318> [Accessed 15 August 2020]

[9] Sparke, L. S.Gallagher, J. S. III (2000). Galaxies in the Universe: An Introduction. Cambridge University Press. ISBN 978-0-521-59740-1. [Accessed 2 October 2020]

[10] Hupp, E.Roy, S.Watzke, M. (August 12, 2006). ”NASA Finds Direct Proof of Dark Matter”. NASA. [Accessed September 20 2020].

[11] Andrew Brock, Jeff Donahue, Karen Simonyan. LARGE SCALE GAN TRAINING FOR HIGH FIDELITY NATURAL IMAGE SYNTHESIS. ICLR 2019. [PDF] Available at: <https://arxiv.org/abs/1809.11096> [Accessed 14 August 2020]

[12] Lizhi Xie, Qi Guo, Andrew P. Cooper, Carlos S. Frenk, Ran Li, Liang Gao, The Size Evolution of Elliptical Galaxie. [PDF] Available at: <https://arxiv.org/abs/1410.2341> [Accessed 2 October 2020]

[13] Seungwook Han, Akash Srivastava, Cole Hurwitz, Prasanna Sattigeri, David D. Cox. not-so-BigGAN: Generating High-Fidelity Images on a Small Compute Budget. [PDF] Available at: <https://arxiv.org/abs/2009.04433> [Accessed 16 September 2020]

[14] Williams, M. J.; Bureau, M.; Cappellari, M. (2010). ”Kinematic constraints on the stellar and dark matter content of spiral and S0 galaxies”. Monthly Notices of the Royal Astronomical Society. 400 (4): 1665–1689. Available at <https://arxiv.org/abs/0909.0680> [Accesed 8 October 2020]

[15] Rumelhart, David E.; Hinton, Geoffrey E.; Williams, Ronald J. (1986a). ”Learning representations by back-propagating errors”. Nature. 323 (6088): 533–536. Bibcode:1986Natur.323..533R. doi:10.1038/323533a0. S2CID 205001834. [Accessed 2 October 2020]

[16] Ivakhnenko, A. G.; Lapa, V. G. (1967). Cybernetics and Forecasting Techniques. American Elsevier Publishing Co. ISBN 978-0-444-00020-0. [Accessed 2 October 2020]

[17] Schmidhuber, Jürgen (2015). ”Deep Learning”. Scholarpedia. 10 (11): 32832. Bibcode:2015SchpJ..1032832S. doi:10.4249/scholarpedia.32832. [Accessed 2 October 2020]

[18] Rina Dechter (1986). Learning while searching in constraint-satisfaction problems. University of California, Computer Science Department, Cognitive Systems Laboratory.Online. [Accessed 2 October 2020]

[19] Mehdi Mirza, Simon Osindero,Conditional Generative Adversarial Nets. [PDF] Available at: <https://arxiv.org/abs/1411.1784> [Accessed 14 September 2020]

[20] Raddick, M. Jordan; Bracey, Georgia; Gay, Pamela L.; Lintott, C. J.; Murray, Phil; Schawinski, Kevin; Szalay, Alexander S.; Vandenberg, Jan (December 2010) [16 September 2009]. ”Galaxy Zoo:Exploring the motivations of citizen science volunteers”. Available at <https://arxiv.org/abs/0909.2925> [Accessed 14 August 2020]

[21] Dor Bank, Noam Koenigstein, Raja Girye , Autoencoders. [PDF] Available at <https://arxiv.org/abs/2003.05991> [Accessed 12 September 2020]

[22] LSST Science Collaboration, P. A. Abell, J. Allison, S. F. Anderson, J. R. Andrew, J. R. P. Angel, L. Armus, D. Arnett, S. J. Asztalos, T. S. Axelrod, and et al. LSST Science Book, Version 2.0. ArXiv e-prints, December 2009.

[23] Alec Radford, Luke Metz, Soumith Chintala, Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. [PDF] Available at <https://arxiv.org/abs/1511.06434>. [Accessed 14 September 2020]

[24] Han Zhang, Tao Xu, Hongsheng Li, Shaoting Zhang, Xiaogang Wang, Xiaolei Huang, Dimitris Metaxas, StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks. [PDF] Available at <https://arxiv.org/abs/1612.03242>, [Accessed 14 September 2020]

[25] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, Sepp Hochreiter, GANs Trained by a Two Time-Scale Update Rule Converge to a

Local Nash Equilibrium. [PDF] Available at <https://arxiv.org/abs/1706.08500>. [Accessed 1 November 2020]

[26] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, Wenzhe Shi, Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. [PDF] Available at <https://arxiv.org/abs/1609.04802> [Accessed 23 October 2020]

[27] Jianpeng Cheng, Li Dong, Mirella Lapata, Long Short-Term Memory-Networks for Machine Reading. [PDF] Available at <https://arxiv.org/abs/1601.06733> [Accessed 1 November 2020

[28] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin, Attention Is All You Need. [PDF] Available at <https://arxiv.org/abs/1706.03762> [Accessed 17 November 2020]

[29] Min Lin, Qiang Chen, Qiang Chen, Network In Network. [PDF] Available at <https://arxiv.org/abs/1312.4400v3> [Accessed 12 March 2021]

[30] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Chen Change Loy, Yu Qiao, Xiaoou Tang, ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. [PDF] Available at <https://arxiv.org/abs/1809.00219> [Accessed 12 March 2021]

[31] Albert Einstein, RELATIVITY THE SPECIAL AND GENERAL THEORY. [PDF] Available at <http://www.f.waseda.jp/sidoli/Einstein_Relativity.pdf> [Accessed 26 March 2021]