FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Echocardiography Automatic Image Quality Enhancement Using Generative Adversarial Networks

Teresa Corado

Mestrado em Engenharia Informática e Computação

Supervisor: João Pedrosa

Second Supervisor: Miguel Tavares Coimbra

July 27, 2023

Echocardiography Automatic Image Quality Enhancement Using Generative Adversarial Networks

Teresa Corado

Mestrado em Engenharia Informática e Computação

July 27, 2023

Abstract

Cardiovascular diseases are known to be responsible for more deaths than any other health issue, and are projected to remain the leading cause of death globally. Cardiac imaging modalities are a key factor to detect several pathologies from which echocardiography can be highlighted. Several upsides of it can be enumerated, such as high temporal resolution, absence of ionizing radiation, its portability and the relatively low costs associated, therefore playing a crucial role in clinical cardiology with diagnostic, prognostic and interventional value.

As such, echocardiography is used as a tool to extract several clinical parameters to evaluate cardiac morphology, deformation and function. Nonetheless, such results require manual delineation and tracking of the cardiac chamber walls or other structures, which can be time consuming for physicians, arising the need for complementary or full automatic analysis tools. Despite the attempts to automate these tasks, the image quality is still a major drawback which decreases its success rate. This happens due to echocardiographic images' low contrast-to-noise ratio, the presence of artefacts and general lower quality due to the dependence on the acquisition conditions and settings. These problems have been previously tackled in other medical areas, namely in the quality enhancement field, mainly resorting to Generative Adversarial Networks.

Thus, the development and validation of automatic image analysis algorithms for image quality enhancement based on image-to-image generative adversarial networks has been conducted, more specifically using cycleGAN, enabling the automatic improvement of image quality.

The dataset used contained two chamber and four chamber views of 500 different patients, comprising a total of 19.240 echocardiography images, along with their respective segmentations. Each sequence was labeled regarding its quality, ranging from low, medium or high quality. A cycleGAN model was developed, aiming to enhance lower quality images and compare them to high quality ones. Two different experiments were conducted. Firstly, the model was trained using low and high quality images. Secondly, low and medium quality images were grouped into one low quality category, therefore using the full dataset during training.

The optimized cycleGAN model was achieved after empirically studying the impact of different variables such as training data and learning rate schedulers. The enhanced images trained with the full dataset in conjunction with the application of a learning rate scheduler achieved the best results, attaining an FID score of 21.539 compared to the baseline of 27.440 with the original dataset echocardiograms. Subsequently, the same model was validated on its impact on an automatic cardiac chamber segmentation algorithm. Its DICE score improved in approximately 14.06% over the baseline after applying the enhanced echocardiography images, improving its automatic segmentation capacities and contributing to an improvement on medical analysis and diagnosis.

The utilization of cycleGAN for echocardiography image enhancement successfully fulfilled its objectives, improving image clarity and quality overall. This enhancement improved the results regarding cardiac chamber automatic segmentations, contributing to the advancement of artificial intelligence in the field of cardiology.

Acknowledgments

I would like to express my gratitude towards my supervisors, Professors João Pedrosa and Miguel Coimbra, for challenging me and providing all the necessary support for the development of this dissertation. A special thank you to Professor João Pedrosa, for the invaluable expertise, patience and guidance provided throughout this journey. I would also like to thank Sofia Ferraz, for all the support and helpfulness through this process.

I am immensely grateful to my friends, for their presence in my life has been an endless source of joy, growth, and support. From listening to my endless rants to staying with me during late-night working sessions, through the highs and the lows, I have had the luck to be surrounded by people who were my mood and confidence boosters, enriching my life in ways words cannot fully express.

Last but certainly not least, I would like to express my heartfelt gratitude towards my family, the foundation of the person I am today. For providing me with everything they could, prioritizing my education and allowing me to grow and pursue my goals. To the ones who stayed and the ones who parted along the way, I know I will always find comfort and guidance in their unwavering love and unyielding support.

Teresa Corado

Contents

List of Figures

List of Tables

Abreviaturas e Símbolos

2D	Two-Dimensional
3D	Three-Dimensional
AI	Artificial Intelligence
CHD	Congenital Heart Disease
CNN	Convolutional Neural Network
CT	Computed Tomography
DL	Deep Learning
DL	Dice Loss
EMD	Earth Mover's Distance
FDG	Fluorodeoxyglucose
FID	Fréchet Inception Distance
GAN	Generative Adversarial Networks
HD	Hausdorff Distance
IS	Inception Score
LV	Left Ventricle
MAD	Mean Absolute Deviation
MRI	Magnetic Resonance Imaging
PET	Positron Emission Tomography
SPECT	Single-Photon Emission Computed Tomography
TTE A4C	Transthoracic Echocardiography Apical 4-Chamber
VTT	Visual Turing Test
WD	Wasserstein Distance

Chapter 1

Introduction

1.1 Context and Objectives

Cardiovascular diseases are the leading cause of death worldwide [25] and echocardiography is a key tool in the detection of pathologies. Despite its advantages, such as high temporal resolution, absence of ionizing radiation and low cost, the quality of images obtained can be poor due to low contrast-to-noise ratio, artefacts and other factors.

This represents an obstacle not only for human operators who could benefit from better image quality, but especially for automatic image interpretation methods such as segmentation and tracking of cardiac chambers.

In order to address this issue, this project aims to develop and validate algorithms for automatic image quality enhancement based on generative adversarial networks (GANs). GANs are a class of deep learning models which consists of two neural networks, the generator and the discriminator, which work in a competitive environment to produce enhanced images. By leveraging the potential of GANs, this project aims to improve the visual quality of echocardiographic images, making them clearer and more informative.

Following the development of the algorithms, various validation techniques will be performed, recurring to evaluation metrics such as the Fréchet Inception Distance (FID). Additionally, the impact of the image quality enhancement algorithms on automatic cardiac chamber segmentation algorithms will be evaluated. By enhancing the images before they undergo segmentation, it is expected to improve the accuracy and reliability of the segmentation process. This could lead to more precise and consistent medical analysis and diagnosis of cardiovascular conditions.

The potential benefits of this research are extensive, as enhanced image quality could enhance the overall performance of echocardiography and therefore contribute to early and accurate detection of cardiovascular diseases.

1.2 Document Structure

This document is structured in six sections, consisting of "Introduction", "Fundamental Concepts", "Image Generation and Enhancement in Echocardiography", "CycleGAN-based Echocardiography Image Quality Enhancement", "Echocardiographic Quality Enhancement in Automatic Segmentation" and "Conclusion and Future Work". The bases and fundamental concepts for this paper are further discriminated in section 2, "Fundamental Concepts", which introduces several cardiology concepts, focusing on cardiovascular imaging techniques, specifically echocardiography. Section 3, "Image Generation and Enhancement in Echocardiography", presents the state of the art, analysing the existing work in similar contexts and the similar approaches and technologies used, followed by the analysis on the conclusions drawn from each of them. In section 4, "CycleGAN-based Echocardiography Image Quality Enhancement", the experimental process regarding the development of the CycleGAN model will be presented along with the results obtained. Subsequently, section 5, "Echocardiographic Quality Enhancement in Automatic Segmentation" aims to analyse the integration of the images obtained from the model developed in section 4 in an automatic segmentation algorithm, explaining the process and its outcomes. Lastly, in section 6, "Conclusion and Future Work", the conclusions of the dissertation are enumerated, as well as potential future work.

Chapter 2

Fundamental Concepts

2.1 The Heart and Cardiovascular Disease

The heart is a vital muscular organ which is responsible for providing the body with oxygenated blood, pumping it to all parts of the body, including the brain and other organs. It is the center of the circulatory system and is structured in four chambers: two atria and two ventricles. It is also constituted by muscle tissue and is responsible for pumping oxygenated blood through the body and deoxygenated blood back to the lungs. Without a healthy heart, the body would not be able to function properly.

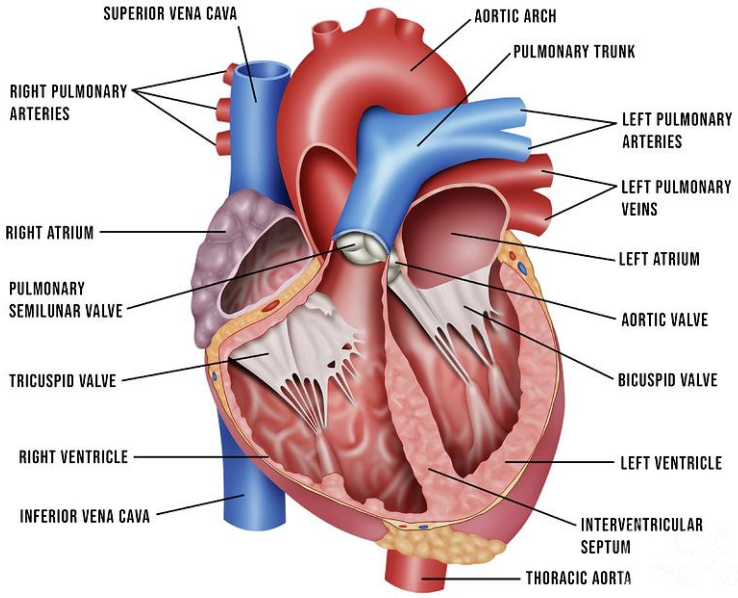

Figure 2.1: Representation of the heart structure [5]

Bearing this in mind, it is important to focus on cardiovascular diseases and their magnitude and impact on the proper functioning of the heart. These are a major and growing cause of death and disability in the world, representing the leading cause of death in most developed countries,

responsible for more than 17.9 [36] million deaths annually. Cardiovascular diseases are also the leading cause of death in low- and middle-income countries, where they are responsible for over 80% [10] of all deaths while its global burden is still increasing, due to aging population, unhealthy lifestyles, and increasing urbanization. These include a range of conditions, such as coronary artery disease, congestive heart failure, and stroke, which can have a major impact on an individual's health, quality of life, and ability to work and participate in society. Therefore, understanding the causes and developing effective treatments for these diseases is of great importance. Early detection and treatment can help reduce the risk of serious complications, including heart attack and stroke, hence the need to improve the existing methods and make them more efficient and accessible.

2.2 Cardiovascular Imaging

It is of substantial importance to understand and analyse the heart's morphology, adapted to each individual, in order to assemble an accurate diagnostic. Such can be achieved through the usage of several different cardiovascular imaging modalities, the main ones being Echocardiogram (echo), Cardiac computed tomography (CT), Single-photon emission computed tomography (SPECT), Cardiac positron emission tomography (PET) and Cardiac Magnetic resonance imaging (MRI).

Echocardiogram (echo) consists in the image resulting from an echocardiography. This modality is widely available and noninvasive, and allows quantitative and qualitative assessment of cardiac anatomy and function, such as heart wall thickness and motion. Additionally, it is also used for assessment of the function and anatomy of the heart valves, detection of valvular vegetations and intracardiac thrombus and providing an approximation of pulmonary arterial pressure and central venous pressure. [33] There are three different echocardiography techniques, these being transthoracic as the most common one, transesophageal and intracardiac.

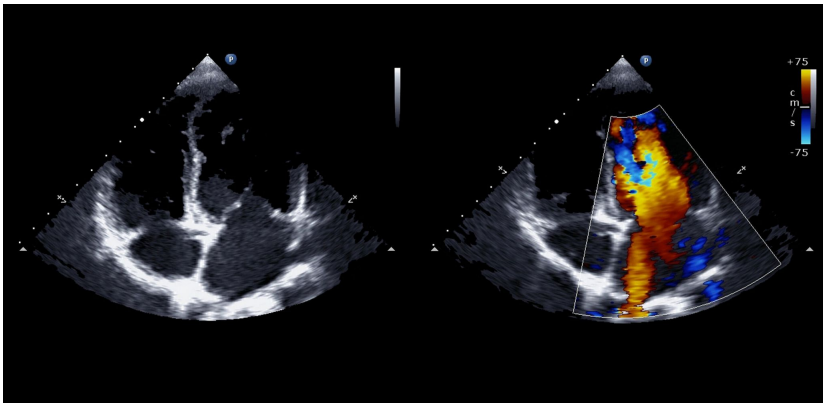

Figure 2.2: Example of an echocardiogram [2]

Among other cardiovascular imaging modalities, **Cardiac computed tomography (CT)** is also widely utilized. In CT, multiple x-ray images are taken from various angles all around the

patient in order to generate two-dimensional images (tomograms) representing a slice of the structure, or three dimensional images of the heart, in addition to great vessels and surrounding structures [19]. CT provides significantly more information compared to conventional x-rays due to a better differentiation between various soft-tissue densities, and its 3D component can be particularly useful in surgery planning. Nevertheless, this modality presents some disadvantages, the main one being a relatively high exposure to radiation, placing the patient at potential risk over time.

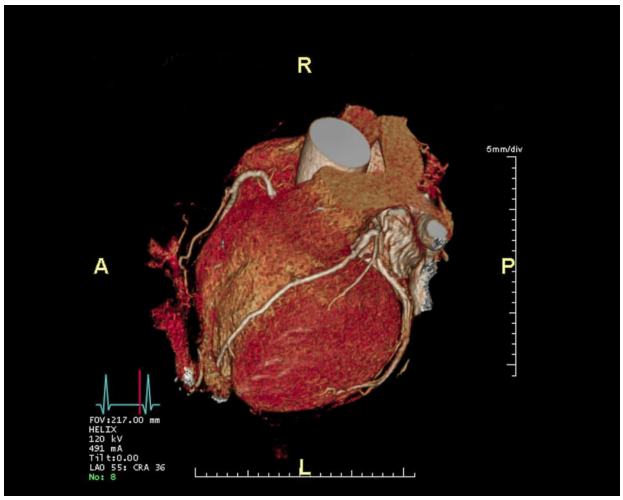

Figure 2.3: Example of a three dimensional cardiac CT [8]

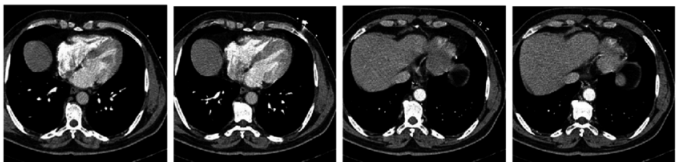

Figure 2.4: Example of a two dimensional cardiac CT [27]

Cardiac Magnetic resonance imaging (MRI) relies on the usage of magnetic fields and radio waves in order to generate images of thin slices of tissues, designated as tomographic images. This method can produce detailed images with high resolution, both two and three dimensional, providing an extensive analysis of heart structure, function and diseases [20]. Although this modality does not emit radiation, it is rather expensive and requires longer imaging times than CT. It may also generate further issues relative to the magnetic field, patient claustrophobia or contrast reactions.

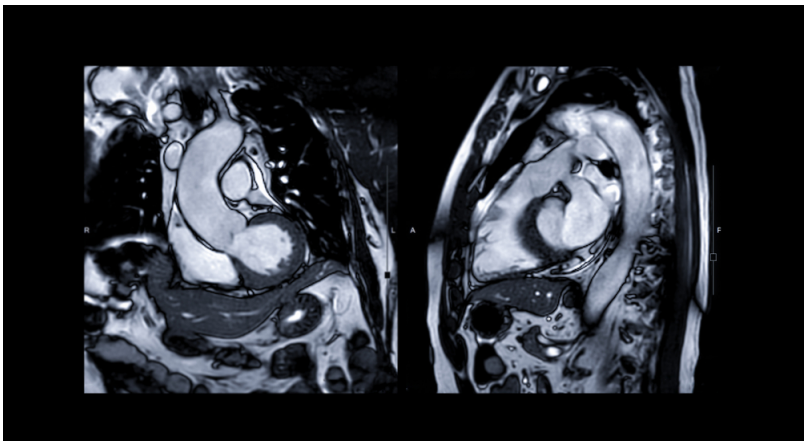

Figure 2.5: Example of a cardiac MRI [7]

Similarly to MRI, **Single-photon emission computed tomography (SPECT)** also allows for three-dimensional cardiac images, however its focus is more leaned to the functioning of the heart such as how the heart is working or how well blood is flowing, rather than the anatomical or structural viewpoint. It consists on a noninvasive method which injects radioactive tracers into a vein, using a rotating camera system and tomographic reconstruction to produce images which may be useful on different diagnostics such as coronary artery disease or the previous occurrence of a heart attack. [1] Some downsides of SPECT involve long scan times as well as lower resolution images susceptible to artifacts and attenuation. Contrarily to **PET**, it also does not provide a quantifiable estimation of blood flow.

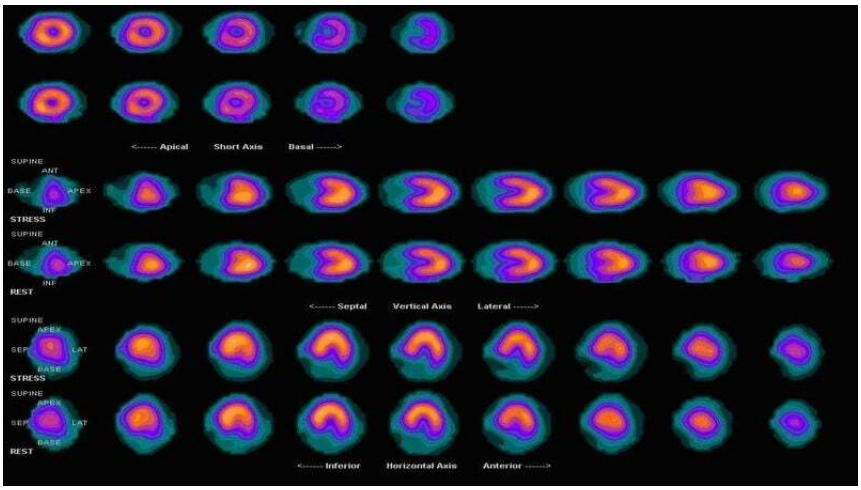

Figure 2.6: Example of SPECT Imaging [26]

Comparably to **SPECT**, **Cardiac positron emission tomography (PET)** consists on a noninvasive nuclear imaging test, using compounds containing radionuclides in order to generate images of the heart. It can provide information regarding tissue function due to the incorporation of positron-emitting radionuclides into metabolically compounds [21]. This can result in several possible applications such as evaluating myocardial viability, diagnosing coronary artery disease

or assessing the damage provoked by a heart attack. Similarly to **SPECT**, it is generally safe, involving a relatively small amount of radiation, however it presents some downsides regarding fluorodeoxyglucose (FDG), which is the most common compound used. Its production requires a cyclotron and has a short half-live, leading to a resulting expense, inconvenience and impracticality which impact the availability of this modality.

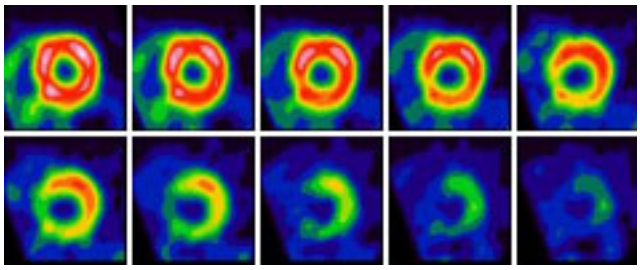

Figure 2.7: Example of a PET Myocardial Perfusion Imaging [35]

While each of the mentioned modalities has its own advantages and is adequate for specific situations, echocardiography can be highlighted since it is widely used due to some characteristics, such as portability, high temporal resolution, absence of radiation, and its low-costs [32]. Additionally, it allows for decision-making with a high degree of accuracy in a multitude of clinical settings. On the other hand, one of the main disadvantages of echocardiography is the relatively low image quality and presence of artifacts whereby a patient might undergo unnecessary other diagnostic tests or interventions, possibly requiring more expenses and risks.

2.2.1 Echocardiography

It is essential to understand the basic principles of ultrasound imaging and Doppler echocardiography in order to achieve accurate results, both during data acquisition and for the interpretation of the ultrasound information. [28] Echocardiography is based on ultrasound waves mechanisms, which consist of mechanical vibrations which can be defined in terms of frequency inducing alternate refraction and compression of any physical medium they pass through.

Sound waves with frequencies ranging from 1.0 to 20 MHz are frequently used in medical ultrasound imaging. The capability of distinguishing separate objects increases directly with frequency and decreases with wavelength. An ultrasound with a high frequency and short wavelength can differentiate objects that are only 1 mm apart, resulting in enhanced spatial resolution. As a result, since the resolution of an echocardiographic image is typically 1 or 2 wavelengths, imaging with a 2.5 MHz transducer would produce an image resolution of about 1 mm. However, due to attenuation, the ability to transmit enough ultrasonic energy into the tissue or the depth of tissue penetration is directly proportional to wavelength and hence inversely related to transducer frequency. Hence, reduced tissue penetration is the trade-off for using higher frequency transducers, and therefore higher spatial resolution.

Each sound wave is a pulse; its frequency is determined by the number of pulses emitted per unit of time, and its "spatial pulse length" consists of the distance from the start to its end. A

pulse moves across a homogenous medium in a straight line. Pulses which are being released collide with tissue surfaces that have various acoustic properties and reflect some of their energy back to the transducer. During the "listening" phase, the echoes that return to the transducer are transformed into an electrical signal which generates a picture on a monitor. [22] However, by changing to a non homogeneous medium, the pulse path is altered, occurring, besides reflection, events of scattering, refraction and attenuation, all three contributing to decreasing the wave's size. Attenuation consists on the gradual loss of energy due to reflection and absorption by conversion to heat, being frequency and wavelength dependent.

2.3 Towards Automated Echo Interpretation

As previously mentioned, automated Echo interpretation plays an important role in improving current imaging modalities, as it can help to enhance the accuracy and speed of diagnosing medical conditions. Automated techniques can have different applications, from quality assessment to full diagnosis, which can provide more accurate results in a fraction of the time it takes a trained doctor to interpret an echocardiogram. Automated echocardiography interpretation can also help to reduce costs associated with diagnosis, as it reduces the need for a trained professional to manually interpret the test results. Furthermore, it can also be used to identify patterns in the data that could not be detected through manual interpretation, as well as to improve patient care by providing more accurate information which can be used to assist on the best possible treatment decisions.

Quality assessment is crucial to ensure the conditions for an accurate automated interpretation diagnosis. Artificial intelligence still presents major drawbacks associated to computer vision in low quality or off standard imaging, as is common for echocardiographic images. Additionally, a better image quality also facilitates the interpretation of a trained professional and can help reducing the time it takes for a full diagnosis. These advantages can be highly expanded through automated image quality enhancement, as quality improvement will inevitably lead to more precise conclusions and faster results.

Chapter 3

Image Generation and Enhancement in Echocardiography

3.1 State of the Art in Image Generation

The rapid adoption of artificial intelligence (AI) technology in various fields has greatly impacted the development of precision medicine, particularly with the growth of medical data and advancements in deep learning technology. One of the most prominent areas of AI application in medicine is in medical image generation.[3] Artificial image generation has been extensively studied and documented due to its many potential applications such as image synthesis, reconstruction, segmentation, noise reduction, and classification [37].

In this section, a comprehensive understanding of the current state of AI image generation will be provided, including the most recent technologies and tools used, as well as their applications. The chapter will delve into the latest developments in AI image generation, including the use of Generative Adversarial Networks (GANs), which have shown promising results in generating high-quality images. The limitations of current AI image generation techniques will also be explored, along with the challenges faced by researchers and practitioners in this field.

3.1.1 Generative Adversarial Networks Fundamentals

Generative Adversarial Networks (GANs) are a technique in machine learning which allows for both semi-supervised and unsupervised learning by modeling high-dimensional data distributions. A GAN consists of two networks: a generator network, G, and a discriminator network, D, as shown in figure 3.1. The generator network creates synthetic images with the goal of making them appear realistic, learning from its interaction with the discriminator network, while the latter receives both the synthetic images and real images and is trained to differentiate between them, classifying the images as real or fake. [3]

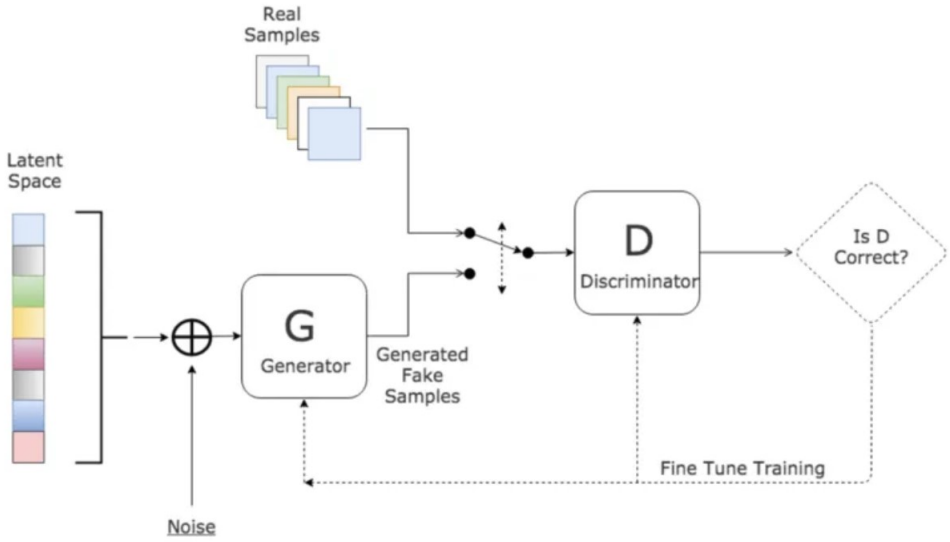

Figure 3.1: Representation of a GAN architecture [11]

The generator and discriminator are trained in parallel and in opposition to one another, through a process known as adversarial training, which consists on updating the weights of both networks based on their respective performance. The generator is trained to reduce the accuracy of the discriminator, while the discriminator is trained to minimize its error in classifying real from fake data, hence if the generator distribution is able to match the real data distribution, the discriminator network will be maximally confused. The end result is that both networks improve over time, and the generator is able to generate increasingly realistic synthetic data.

GANs have a number of key applications, including image synthesis, image-to-image translation, and style transfer. For instance, GANs can be trained on a dataset of real images in order to produce synthetic images which resemble the real images, although with some variation. This has applications in various fields such as medical sciences, where synthetic images can be used to generate training data for other machine learning algorithms.

However, a challenging aspect of GANs is that they can be difficult to train, due to the competing objectives of the generator and discriminator. This can result in a phenomenon known as mode collapse, where the generator fails to generate a diverse range of synthetic data. There are several strategies for overcoming this challenge, including the use of a variety of GAN architectures, and the use of techniques such as progressive growing and self-attention.

3.1.2 Image to Image GANs

3.1.2.1 Pix2pix

The pix2pix model is based on conditional adversarial networks and, as an image-to-image GAN, it is designed to translate an input image into an output image. This is a common task in computer graphics, image processing, and computer vision. [3] The model is trained by learning the

mapping between the input image and the output image and constructing a loss function to optimize the mapping. More specifically, it uses a U-Net [30] generator network and a Convolutional PatchGAN discriminator network to translate an input image into an output image. After that, the model is trained by optimizing the CGAN loss function with an additional L1 distance term to gauge how close the produced and real images are. This produces a generated image with less blurring. In order to prevent vanishing gradients [15], the generator network is trained to maximize $log(D(x,G(x,z)))$, while the hyper-parameter regulates the weight of the L1 distance term. The Pix2Pix algorithm's requirement for paired photos, which can be challenging or expensive to get, is one of its primary flaws.

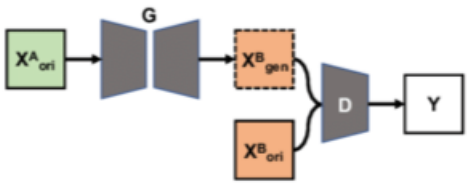

Figure 3.2: Representation of a pix2pix architecture [18]

3.1.2.2 CycleGAN

Alternatively to the pix2pix model, CycleGAN is an unsupervised generative model which does not require paired images. As long as there is enough variety in the source and target data, it can be applied to images that are not associated. Two cycles make up the model, which is divided into four networks. As illustrated in figure 3.3, two of these networks are generator networks (G1 and G2), and the other two are discriminator networks (D1 and D2) [38].

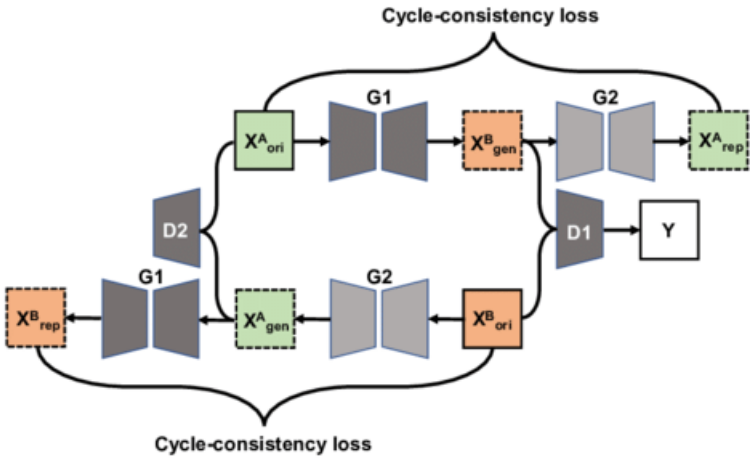

Figure 3.3: Representation of a CycleGAN architecture [18]

The generators and discriminators are trained in an adversarial manner, where the generators aim to produce images which are indistinguishable from real images, while the discriminators aim

to distinguish between real and generated images. During training, the generators and discriminators are updated alternately, with both being updated in order to improve their abilities to fulfill each objective.

One key feature of the CycleGAN is the use of cycle consistency loss, which ensures that the translated images are consistent with their original inputs. This is achieved by comparing the original image with the translated image and then translating it back to its original domain using the remaining generator. The cycle consistency loss ensures that the original image and the translated-back image are similar, which encourages the generators to produce semantically meaningful translations.

3.2 Evaluation Methods

Due to the lack of a single metric which may capture the various features of image quality, such as realism, diversity, and stability, evaluating GANs is frequently a difficult endeavor. The quality and similarity of created images to genuine images is not measured by the objective function of the generator and discriminator networks, which simply compares how well these networks perform against one another. Therefore, in order to address this issue, a combination of both qualitative and quantitative evaluation measures is needed. In this section, various methods used to evaluate the performance of GANs will be discussed.

3.2.1 Qualitative Measures

Perceptual studies represent one of the most common methods for evaluating the legitimacy of generated samples in GANs, although prone to subjectivity. In this method, human evaluators are presented with real and generated samples and requested to rate them as real or fake, quantifying the model's quality. Human evaluation is the most comprehensive evaluation method as it takes into account all aspects of image quality, including realism, diversity, and stability. However, it is also the most time-consuming and expensive evaluation method, besides being exposed to subjective factors such as the setup, chosen samples or the annotators themselves.

The Visual Turing Test (VTT) is another example of a qualitative evaluation method applied to GANs in order to assess the quality of generated images [12]. This method is based on the concept of the Turing Test, which consists on the measure of a machine's ability to exhibit intelligent behavior equivalent to, or indistinguishable from, that of a human.

In the case of the VTT, a human evaluator is presented with a set of images, some of which are real and others generated by the GAN. The evaluator is then asked a series of binary questions regarding the set of images, providing a qualitative evaluation of the quality of the generated images and the system's ability to recognize objects and attributes, as well as relationships between them. One advantage of the VTT is that it provides a more faithful evaluation of the generated images, which is of high importance in certain applications where the quality of the generated images is critical, such as in medical imaging. On the other hand, the VTT is inherently subjective and may be affected by individual biases and limitations in human perception.

3.2.2 Quantitative Measures

The most common way to evaluate GANs in a quantitative way aims to compare the generated outputs to real-world data. This method involves computing metrics such as the Inception Score (IS) or the Fréchet Inception Distance (FID), which measure the quality and similarity of the generated images to real images.

The IS aims to measure the quality and diversity of generated images by combining two factors: the marginal likelihood of generated images and the entropy of the predicted class probabilities. The goal of this metric is to measure the performance of a generative model by evaluating the generated samples in terms of their perceived quality and diversity. A high Inception Score indicates that the generated images are both high-quality and diverse, whereas a low score suggests that the images are either low-quality or have low diversity.

In order to calculate the Inception Score, the generated images are first passed through an Inception Network, a pre-trained convolutional neural network (CNN) for image classification [31]. The Inception Network produces a predicted class probability for each image, and the Inception Score is calculated as the exponential of the average entropy of these predicted class probabilities across all generated images.

$$IS = \exp(\mathbb{E}_{x\sim p_g} D_{KL}(p(y|x)\|p(y))) \tag{3.1}$$

In the equation 3.1, x consists of a generated image, $p(y|x)$ represents the conditional probability of the class label y given the image x, and $p(y)$ is the marginal probability of the class label y. The KL divergence between $p(y|x)$ and $p(y)$ measures the difference between the class distributions in the generated dataset and the real dataset.

As previously mentioned, the FID is another widely used evaluation metric for generative models, particularly GANs. Similarly to the Inception Score, it aims to measure the quality and diversity of generated images, however it does so in a more direct and detailed manner.

The FID measures the distance between the feature representations of real and generated images in a feature space defined by a pre-trained convolutional neural network (CNN), such as the Inception Network [13]. The feature representations are calculated as the activations of a particular layer of the CNN, and they capture the high-level semantic information regarding the images.

The distance between the real and generated feature representations is calculated using the Fréchet distance, which measures the similarity between two multivariate probability distributions, as represented in the equation 3.2. Regarding FID, the distributions consist on the distributions of feature activations for the real and generated images.

A low FID score indicates that the generated images are similar in terms of their semantic content to the real images, whereas a high score indicates a large difference between the real and generated images.

$$FID(G,R) = \|\mu_g - \mu_r\|^2 + Tr(\mathbf{C}_g + \mathbf{C}_r - 2(\mathbf{C}_g\mathbf{C}_r)^{\frac{1}{2}}) \tag{3.2}$$

In the FID equation, G and R denote the generated images and real images, respectively, μ_g and μ_r are the mean activations of the Inception network for the generated and real images, and $\mathbf{C}_g$ and $\mathbf{C}_r$ consist of the covariance matrices of the activations.

The FID therefore provides a detailed evaluation of the quality and diversity of generated images by comparing the semantic content of real and generated images in a feature space defined by a pre-trained CNN. It is considered to be a more reliable and direct metric compared to the Inception Score, but it is also more computationally expensive to compute.

In cases where real-world data is not available or not appropriate for evaluation, non-data-based methods can be used. For instance, the Frechet Distance (FD) measures the similarity between two distributions, without requiring real-world data. The Wasserstein Distance (WD) and the Earth Mover's Distance (EMD) are other methods which evaluate the quality of the generated images by comparing them to a noise distribution or to a target distribution, respectively.

3.3 Applications in Echo Image Enhancement

Automatic image quality enhancement algorithms applied to medical fields have been previously and continuously explored. In the present chapter, previous works on this field will be introduced and discussed on their results and relevance to this dissertation.

3.3.1 State of the art

Automatic image quality enhancement using AI in the medical field can be particularly helpful and time saving when performing an examination and diagnosis. This topic has been previously explored, namely in the echocardiography area as shown in table 3.1.

For instance, Maria Escobar et al. [9] sought to develop "UltraGAN", a method for ultrasound enhancement which is able to transfer quality details while also preserving structural information of the heart and therefore maintaining anatomical consistency. This project started by using the CAMUS dataset [23] where each echocardiogram was labeled into three categories: high, medium or low quality. The development started out by reducing this into only 2 categories, considering all medium quality images as low quality, using 80% of the images for training and 20% for testing. A training on a simple U-Net model for LV segmentation was performed, using cross validation on the dataset images. Afterwards, an UltraGAN was used in order to improve the quality of the training images for data augmentation, and the U-Net was trained again by merging the two sets of images.

The architecture of the UltraGAN consists of two main components: a generator network and a discriminator network, as shown in figure 3.4.

Authors	Year	Model used	Dataset	Inclusion of temporal information	Results
M. Escobar et al.	2020	UltraGAN	CAMUS	No	Average Dice Score of 89.6%
Jafari, M. et al.	2020	Fully convolutional deep translation model	1089 echo studies from 841 patients	No	Improvement of 30% in the LV segmentation Dice score and 34mm in Hausdorff distance metrics
Jafari, M. et al.	2019	ACCGAN	854 annotated AP4 frames	No	Worst-case Dice score improved 15% over the baseline
Liao et al.	2019	QT-StarGAN	16,612 echo cine series from 3,157 patients	Yes	Improved classification accuracy above 60% user-defined quality level for DenseNet and DenseNet+LSTM, and at 80% for VGG-16
Diller et al.	2019	Custom Autoencoder Network	267 subjects with routine transthoracic examinations	No	$p < 0.005$

Table 3.1: Comparative analysis between the papers presented

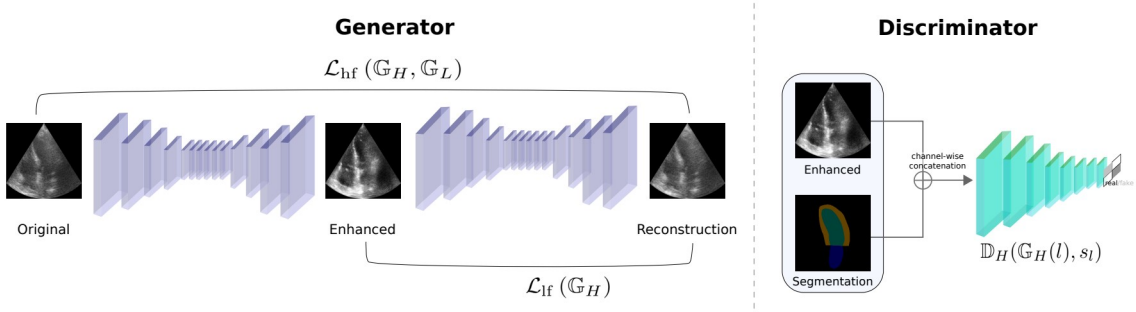

Figure 3.4: UtraGAN architecture [9]

The generator network is designed to enhance low-quality ultrasound images and is based on a U-Net architecture. The U-Net architecture consists of an encoder part which down-samples the input image, followed by a decoder part that up-samples the down-sampled feature maps to produce the final output image. Skip connections are used to concatenate the feature maps from the corresponding encoder and decoder parts, allowing the network to capture both low- and high-level information from the input image.

The discriminator network, on the other hand, is designed to distinguish between real high-quality ultrasound images and fake images generated by the generator network. It is based on

a patch-based discriminator, where the input image is divided into multiple overlapping patches, and each patch is fed into a separate branch of the network. The outputs from all branches are concatenated and passed through a fully connected layer in order to produce the final classification. To ensure that the generated results are anatomically accurate, the discriminator takes two inputs - the ultrasound image (whether it is real or generated) and the segmentation of the relevant corresponding structures.

During training, the generator and discriminator networks are optimized using the adversarial loss, which consists of a combination of the binary cross-entropy loss and a gradient penalty term. The binary cross-entropy loss measures the difference between the predicted and ground-truth labels for real and fake images, while the gradient penalty term helps to ensure that the generated images have a similar distribution to the real images.

Results show that increasing the training data improved the segmentation results in practically all images, suggesting that it correctly preserves the anatomical structure of the images, which is crucial in the context of health diagnosis. The UltraGAN is therefore able to learn a mapping from low-quality to high-quality ultrasound images and produce enhanced images which are more visually appealing and useful for medical diagnosis.

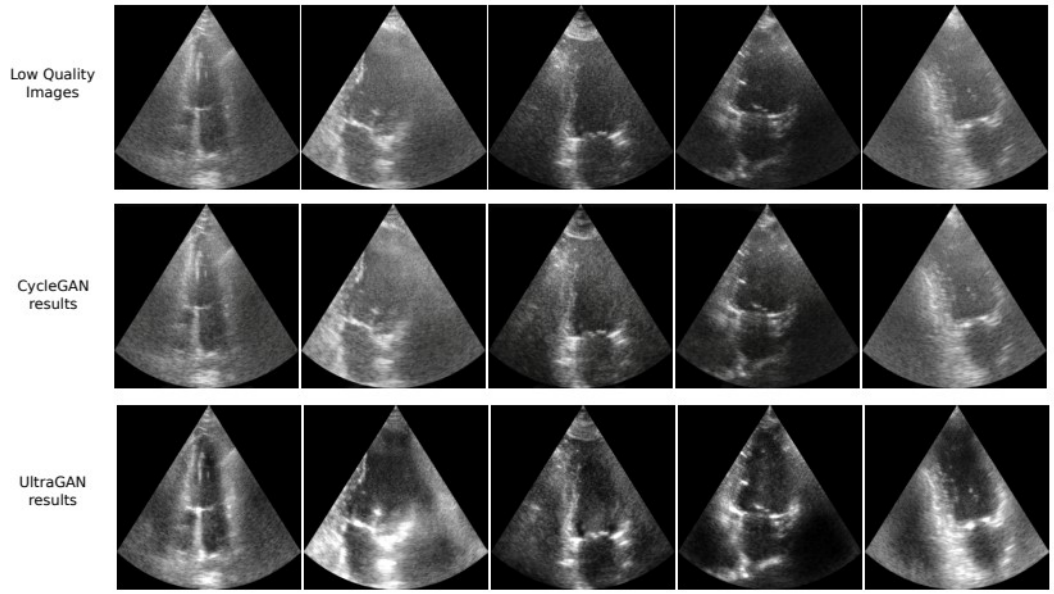

Figure 3.5: Qualitative results obtained by UltraGAN compared to CycleGAN [9]

Furthermore, Jafari, M. et al. [17] also proposed the use of GANs in order to perform echocardiographic image enhancement. More specifically, an anatomically restricted CycleGAN (ACCGAN) was proposed in order to improve the quality of echocardiography in the A4C view (4-chamber), for LV segmentation purposes.

The ACCGAN (Anatomically Constrained CycleGAN) architecture is a variant of the CycleGAN which takes into account the anatomical structure of the echocardiography images during

training. As a GAN, it consists mainly of a generator network and a discriminator network, as represented in figure 3.6.

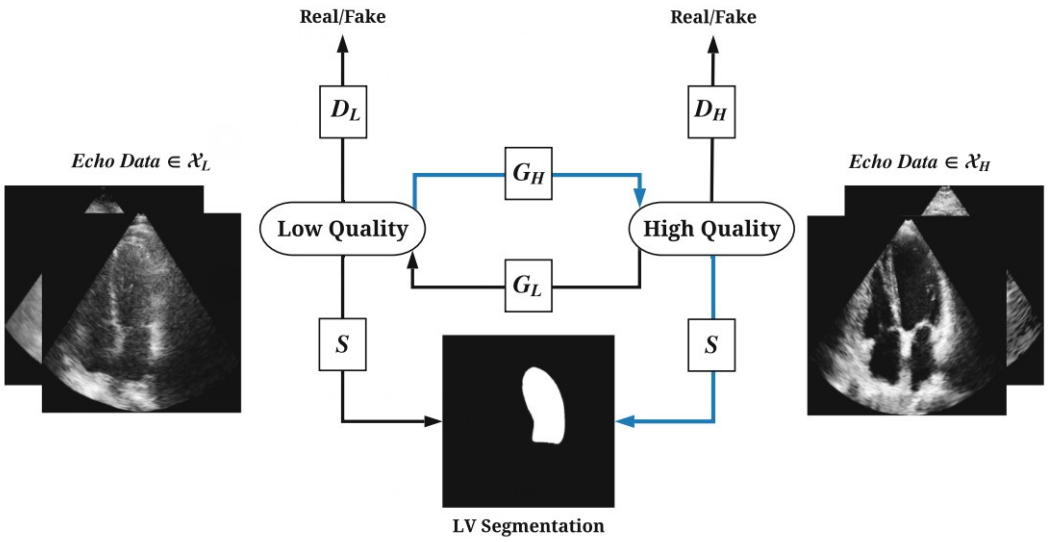

Figure 3.6: Representation of the ACCGAN architecture [17]

A dataset of 854 images with annotations regarding their quality was used and the results showed that the proposed method improved the robustness of the LV segmentation, with the worst-case Dice similarity score increasing by 15% over the baseline. Visual results, which can be analysed in figure 3.7, demonstrated that the ACCGAN network learned to distinguish between high-quality and low-quality echocardiography images and improved the visibility and anatomy of LV in the high-quality images. Additionally, the proposed method is faster and results in a more accurate LV segmentation comparatively to CycleGAN.

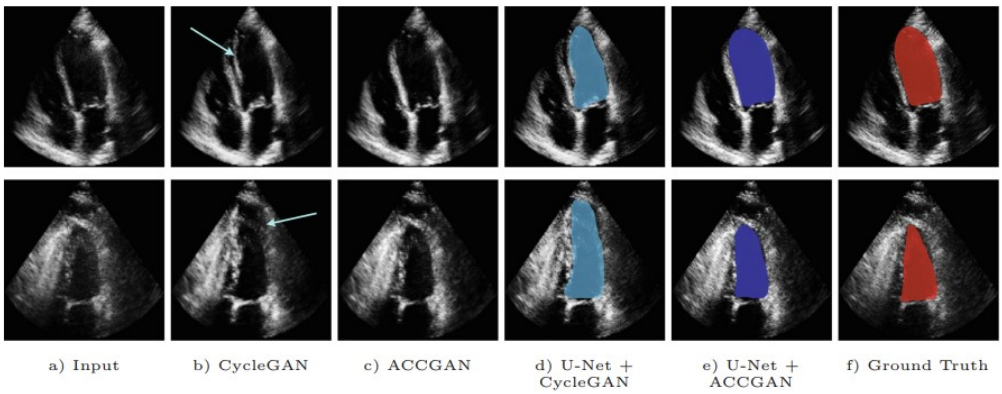

Figure 3.7: Qualitative results obtained by ACCGAN in automatic segmentation algorithms compared to CycleGAN [17]

Furthermore, the same group [16] later attempted to improve the quality of POCUS images to match that of high-end cart-based US systems through a fully convolutional deep translation

model. The model was trained using a constrained cycle-consistent GAN and the proposed system resulted in improved accuracy of LV segmentation from an apical cardiac view.

Analogously, Liao et al. [24] proposed a quality transfer network for echocardiography images, called QT-StarGAN, which uses the StarGAN approach to translate images to a desired quality level. Unlike the previous papers, QT-StarGAN incorporates temporal information during training.

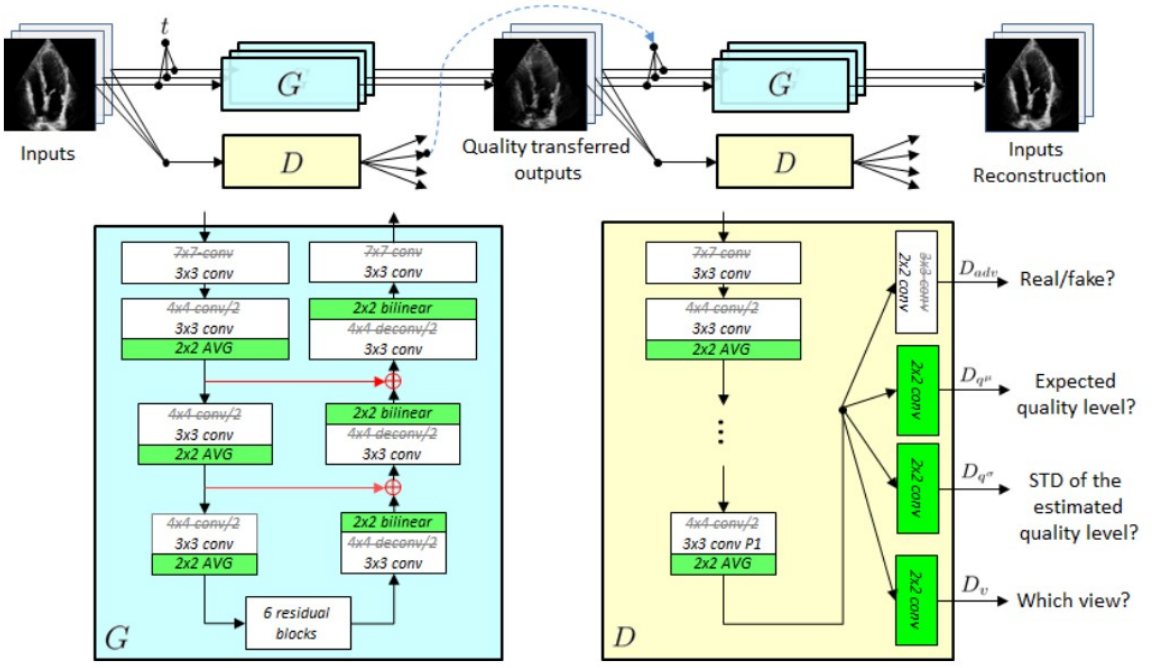

Figure 3.8: Representation of the QT-StarGAN architecture [24]

In the original StarGAN network, the generator is in the form of an auto-encoder with a series of down-sampling convolution layers (contraction path), a series of residual layers (feature tuning path), and a series of up-sampling deconvolution layers (expansion path). The discriminator is a multi-layer convolutional network. In order to improve the performance of StarGAN, the authors made three main modifications to the architecture. Firstly, the deconvolution layers were replaced with bilinear up-sampling layers, so as to reduce visually noticeable artifacts. Additionally, skip connections between the contraction and expansion paths were added in order to allow for the utilization of high-frequency information in image generation. Lastly, the temporal window of the discriminator was extended to three frames, in order to allow for the utilization of temporal information in optimization. These modifications are illustrated in figure 3.8 and examples of their effects are shown in 3.9.

The approach was evaluated with 16,612 echo images from 3,157 patients, showing improved accuracy in a standard echo view classification task using QT-StarGAN.

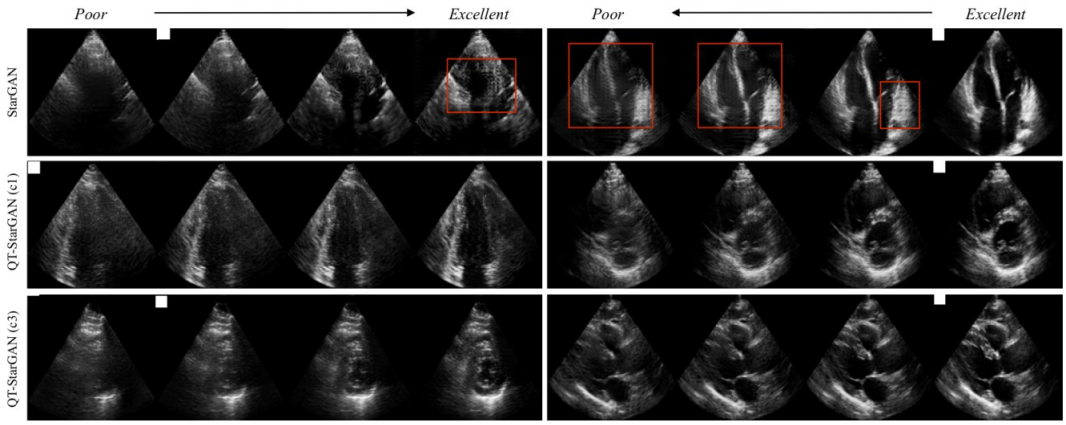

Figure 3.9: Qualitative results obtained by QT-StarGAN [24]

Additionally, Diller et al. [6] aimed to explore the effectiveness of autoencoders, a type of deep neural network, in eliminating noise and removing acoustic shadowing artifacts in Transthoracic Echocardiography Apical 4-Chamber (TTE A4C) views, particularly in patients with Congenital Heart Disease (CHD).

The autoencoder consisted of an encoder, a hidden coding layer, and a decoder that reconstructed the original image from a modified version of the image with added noise, as represented in figure 3.10. The performance of the autoencoder was measured by binary cross-entropy and sum of squared differences, and visually assessed by two experienced investigators. The autoencoder was trained using an adadelta optimization, binary cross-entropy as the loss function, and a batch size of 16 over 200-400 epochs.

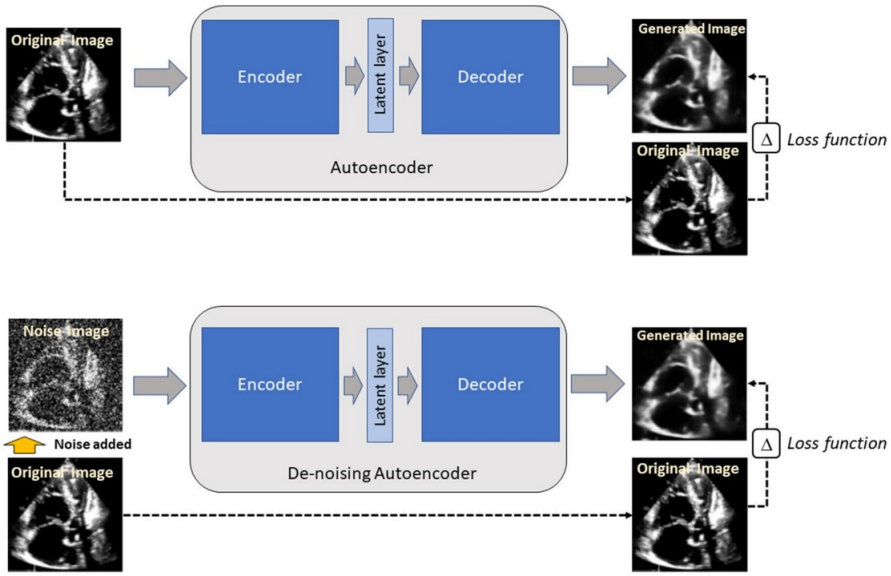

Figure 3.10: General architecture of an autoencoder [6]

The deep learning algorithms specifically designed for CHD samples were compared to those trained only on structurally normal heart samples. Results showed that the proposed network significantly improved the image quality in all diagnostic subgroups ($p < 0.005$) and that the models trained on CHD samples performed better when applied to CHD patient examples.

3.3.2 Datasets

The CAMUS [23] dataset, integrated in the CAMUS project, is the largest publicly-available and fully-annotated dataset for 2D echocardiographic assessment. This dataset contains clinical exams from 500 patients, collected at the University Hospital of St Etienne (France) in accordance with the local ethical committee regulations. The exams were designed to measure left ventricle ejection fraction and no pre-selection or data selection was done in order to ensure a realistic clinical representation. The dataset covers a wide range of acquisition settings and for some patients, parts of the wall were not visible on the images, which might be useful when detecting poorer quality images. For certain cases, the recommended four-chamber view could not be obtained, so a five-chamber view was used instead. This results in a highly heterogeneous dataset, typical of daily clinical practice, including both image quality and pathological cases.

The dataset incorporates a training set of 450 patients with corresponding manual references based on the analysis of a clinical expert, plus a testing set of 50 other patients, where the raw input images are provided in the raw/mhd file format. The quality annotations are divided into three categories, these being "Poor", "Medium" or "Good".

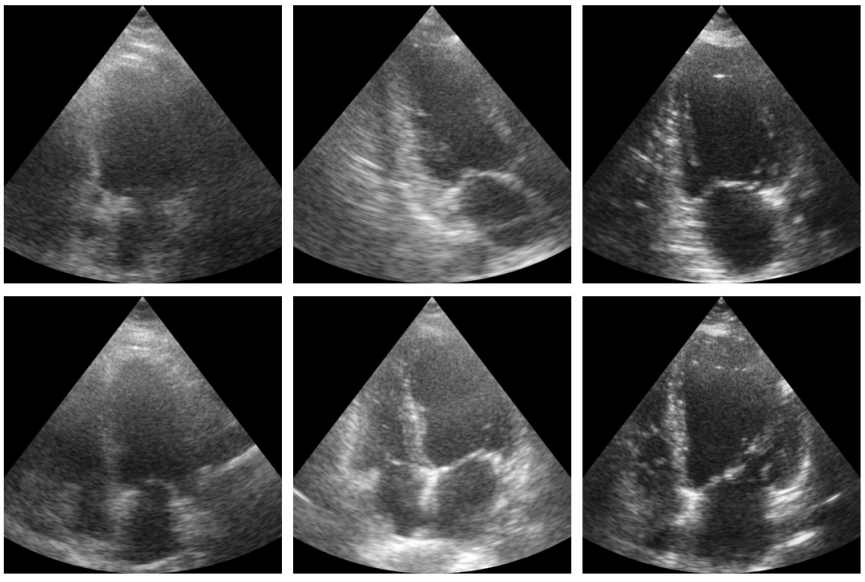

Figure 3.11: Images from the CAMUS dataset, from "poor" to "good" quality, the first row representing the two chamber view and the bottom row the four chamber view of the same patient [23]

Chapter 4

CycleGAN-based Echocardiography Image Quality Enhancement

As previously highlighted, echocardiography is a diagnostic tool of extreme importance, being vital in the field of cardiovascular medicine. It provides real-time imaging of the heart's structure and functioning, however, the image quality remains a major drawback in its development. Echocardiography images are often compromised by various aspects such as artifacts, low contrast-to-noise ratio and suboptimal image acquisition techniques, resulting in general lower quality. These limitations can lead to inaccurate diagnosis, resulting in potential errors in patient treatment and care.

This chapter aims to present the implementation of an approach for echocardiography image quality enhancement using a CycleGAN-based framework. The presented method represents a powerful tool in this context due to the CycleGAN structure and functionality. As an approach to deep convolutional neural networks effective for image-to-image translation tasks, cycleGAN offers a significant potential for enhancing lower quality echocardiography images while receiving non paired training data.

The following sections will provide a comprehensive understanding of the experimental setup, including details regarding the dataset used, the architecture of the CycleGAN model, and the training procedure employed, as well as the optimization techniques applied. The rationale behind these choices will be discussed and the evaluation metrics used to assess the performance of the model will be further presented, as well as the obtained results.

4.1 Methodology

4.1.1 Dataset Overview

The dataset used consists of the CAMUS [23] dataset, which contains echocardiographic images and the respective segmentations of the left ventricle, left ventricular myocardium and left atrium of 500 patients, each with annotations as to their quality, established by cardiologists. These images have been divided into training and testing sets, containing the echos of 450 and 50 patients

respectively, each patient having two image sequences associated - a two chamber view and a four chamber view. Since each view is composed by a sequence of around 18 images, it results in a total of 17,274 training images and 1,968 for testing.

Regarding the final data partitioning, four distinct subsets were used. Both the training and testing data were divided into high quality images and a gathering of lower quality images. The partition also took into consideration the distribution of the patients across the training and testing sets, avoiding the presence of the same patient in both sets. The number of data contained in each set is presented in the table 4.1.

	Training	**Testing**	**Total**
High Quality	8,949	950	9,899
Medium Quality	6,367	731	7,098
Low Quality	1,956	287	2,243
Total	17,272	1,968	19,240

Table 4.1: Dataset Distribution

4.1.2 Model Architecture

This section provides a thorough overview of the model's architecture, as well as its application to the image-to-image translation task. The model's configuration and training details, along with the modifications and enhancements made to the original design to increase its efficiency will be discussed.

Preprocessing of Input Images

The original dataset images were preprocessed in order to better fit the model's requirements and achieve better results. Due to the specificity of the data, the images were firstly resized to a squared shape in order to later be introduced into the model. The pictures were resized without getting deformed by cutting either the left and right sides or the bottom part, since it prevented the relevant data from being lost. This process aimed to preserve an isotropic pixel size and took into consideration the relatively low information present in the bottom and sides of echocardiograms. Given the computational limitations, during training, all images were resized to a fixed resolution of 256 x 256 pixels, in order to establish a balance between data preservation and computational efficiency. These dimensions additionally allowed for all the data to be uniform and consistent throughout the entirety of the training process.

Model Specifications

CycleGAN consists of a deep learning architecture designed for unsupervised image-to-image translation tasks. Without paired training data, it uses the capacity of generative adversarial networks (GANs) to establish the mapping between two separate picture domains. CycleGAN, in

contrast to conventional GANs, does not need one-to-one association between pictures, therefore making it fit for situations where acquiring paired data is difficult or impracticable. Due to the lack of paired data regarding the quality of echocardiography images, CycleGAN is a suitable application with significant potential in this context. Therefore, the main goal of its application in echocardiography image quality enhancement is the improvement of visual understanding, noise reduction and overall improvement of diagnostic quality. This can be achieved due to the resulting mapping between the lower quality images and their corresponding high quality equivalents. From this mapping, the trained model can be used to generate the high quality counterparts and thereby assisting in a more accurate diagnosis.

The model's training process consists of two main steps, these being the adversarial training and the cycle-consistency training. During the adversarial training phase, the generator and discriminator networks are set against one another in a competitive environment. While the discriminator's goal is to differentiate between generated and original images, the generator seeks to improve low quality echocardiography images into higher quality ones. Both networks constantly enhance their performance through backpropagation and gradient descent.

CycleGAN also employs cycle consistency as pictured in figure 4.1 to guarantee that the generated images are consistent with the input domain. As a result, an echocardiogram should not change when enhanced from low to high-quality domain and back again. This restriction is enforced during training by the cycle-consistency loss, which enables the model to reliably learn the mapping functions. This constraint is particularly important in the current context due to the specificity and sensitivity of the data.

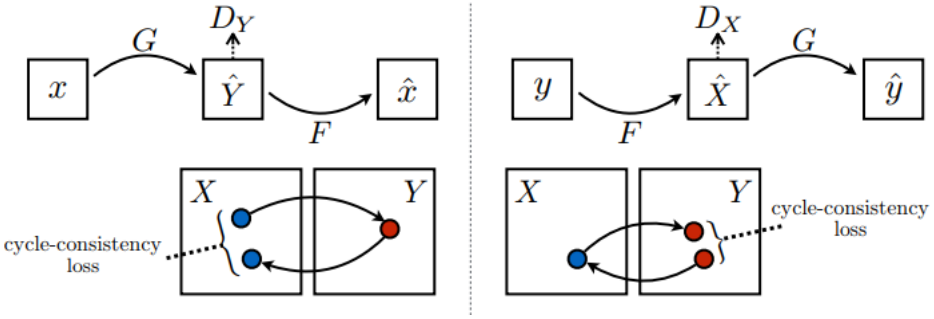

Figure 4.1: CycleGAN consistency loss diagram [4]

Moreover, the generator and discriminator losses have been calculated by recurring to Mean Squared Error Loss (MSE), a commonly used loss function in deep learning tasks which calculates the mean squared difference between the predicted output and the target output.

The generator loss, in this context, is calculated through the MSE loss function. In order to do so, the generated image is fed into the discriminator network, which produces a prediction based on this generated image. This prediction is then compared to the expected output for real images. Therefore, the generator loss is calculated by quantifying the discrepancy between the discriminator's assessment of the generated image and the expected output for real images.

The overall generator loss is determined by taking the average of the losses obtained from evaluating the generated images in both directions of the GAN (from the low quality domain to high quality domain and vice versa). This average represents the generator's ability to generate realistic images that can deceive the discriminator.

Similarly, for fake images produced by the generator, the discriminator's prediction is compared to the expected "fake" label while also using the MSE loss function, which aims to quantify the difference between the discriminator's evaluation and the desired classification. Its loss is computed by combining the losses obtained from real and fake images. The main objective is to minimize this loss, enabling the discriminator to accurately classify real and fake images.

Conversely, the generator aims to maximize the discriminator loss. By doing so, the generator strives to generate fake images which are classified as real by the discriminator, ultimately improving its ability to produce more convincing and realistic images.

Through this adversarial process, the generator and discriminator update their parameters in an iterative way, each trying to outperform the other. This adversarial interplay drives the GAN towards generating reliable high quality images.

4.2 Experiments

This chapter aims to delve into the results obtained by the developed model and to provide a methodical assessment of its performance in various aspects. Through a meticulous examination of the results, added to the architectural details previously mentioned, this analysis intents to offer insights into the performance and constraints of CycleGAN for echocardiographic image translation tasks. Two main configuration factors will be distinguished, these being the influence of both the training data and the application of a learning rate scheduler.

The experimental process consisted on the CycleGAN training, followed by the application of the model in image generation. The image generation step consisted on the generation of high quality images from the lower quality images on the testing set by applying the resulting models from the various epochs. For each training and for each epoch, all lower quality testing images were converted to high quality ones using the corresponding model. The generated data then served as input to calculate the FID score, a widely adopted metric used in quality assessment of images generated by a specific model.

4.2.1 Influence of training data

In this section, the influence of the training data will be studied along with its impact in the model's development and evaluation results. Two different training options will be presented, based on the variation of the usage of medium quality images during training.

Model 1 - Low and High Quality Based Training

As previously stated, each image sequence from the dataset is labeled according its quality, this being low, medium or high. Initially, only high and low quality images were considered, in order for the distinction to be clearer for the algorithm. This option significantly diminished the amount of data available, since medium quality images were not considered neither for training nor for image generation purposes. This option resulted in a reduction from 17,272 training images and 1,968 testing images to a total of 10,905 and 1,237, respectively.

Therefore, in the initial experiments, only low quality images were converted by the model to high quality ones, which resulted in 287 generated images per epoch.

In order to establish the baseline, the FID score was firstly calculated with the original images, comparing the high quality testing images to the low quality ones, without any modifications, representing the control group. This step allowed for a more accurate comparison between the generated images and the ground truth and its dissimilarities.

Subsequently, the FID score was calculated for each set of generated images. Through the comparison between the FID scores of the generated images and the control value, it was possible to objectively measure the performance of the different models in enhancing the original images. This procedure allowed for a meticulous assessment of the model's image enhancement capabilities and provided valuable insights into the extent to which it was able to maintain the quality and integrity of the data.

Afterwards, the model was tested on the full testing set, this including both medium and low quality images, increasing its size to 1018 generated images. Therefore, despite having been trained using only low quality images, its improvement capacity of medium quality images was tested. Similarly to the previous experiment, the FID baseline was calculated comparing low and medium quality images to high quiality ones.

Model 2 - Low, Medium and High Quality Based Training

Due to the reduced training size of the previous experiment, as an alternative experimental setting and similarly to Maria Escobar et al. [9], low and medium quality images were grouped into a single low quality category. This decision took into account the cycleGAN structure regarding the usage of two different sets of data for model training, added to the intention of leveraging the entirety of the dataset. Thus, a larger sample of data was used, containing both low and medium quality images from the training set, which allowed for a usage of the totality of the dataset.

Similarly to the first experiment, this model underwent both FID evaluation scenarios of generated images from the testing set. Firstly, only low quality images were enhanced and its FID calculated, followed by the enhancement of the gathering of both low and medium quality images and their respective FID scores. This method allowed for a more precise comparison between both models in order to assess the impact of the training data in model performance.

4.2.2 Influence of learning rate scheduler

The CycleGAN model architecture comprises two fundamental components, the generator and discriminator networks. Within the context of the current problem, the generator network aims to generate both high quality images from low quality inputs and the reverse transformation, whereas the discriminator is responsible for distinguishing between real and generated images. In order to ensure continuous improvement of the model, it is crucial to establish a balance in the competitive dynamics between the generator and discriminator networks. This balance drives their progressive enhancement as they mutually challenge each other to refine their respective performances. The learning rate of each network consists on a hyper-parameter which tunes the learning evolution of the network, and therefore is crucial to maintaining this balance. The tuning of the learning rate scheduler is widely used as a tool to balance the training of the model. As a result, the influence of the application of a learning rate scheduler with different values has been studied and thoroughly analysed in the following section.

Model Variations

In the initial trainings, the same base learning rate of 0.0001 was used for both networks, as an experimental approach. This experimental setup allowed for an exploration on how the discriminator and generator interacted under the same learning rate setting. Notably, such approach has the potential to yield valuable insights into the learning dynamics and the potential imbalance between the two networks.

In order to further explore the learning rate dynamic and its impact on each network throughout training, a learning rate scheduler was incorporated to tune each network's evolution individually and therefore optimize the training process. The learning rate scheduler was adapted to each network's characteristics and dynamically manipulated the learning rates, allowing the model to converge more efficiently. The determination of its values was determined empirically through rigorous experimentation, enabling the identification of ideal settings for an optimized training performance.

Since the learning rate diminished throughout the training process, the base learning rate was firstly increased from 0.0001 and set to 0.0002 for both networks. Each network's learning rate then varied according to fixed factors after a specified number of epochs, which was set to 5. This means that, from the decaying epoch number 5, each network's learning rate decreased constantly after each 5 epochs. This approach was expected to allow for more stable and accurate updates to the model's weights, and consequently improving the loss graphs and FID scores.

Therefore, four different model versions were made in order to empirically assess the influence of different learning rate values during the training process and consequent model performance. All models were trained grouping both low and medium quality into one, similarly to Model 2. Throughout all versions, the discriminator decaying factor was higher, meaning that its learning rate decreased faster than the generator's. The decaying factors for each network of each model

are of the following type: $N_1^{\left(\frac{epoch}{decay\_epoch}\right)} \times N_2^{\left\lfloor\frac{epoch}{decay\_epoch}\right\rfloor}$. The N1 and N2 values are specified for each model's network in table 4.2

	Generator N1	**Generator N2**	**Discriminator N1**	**Discriminator N2**
Model 3	0.95	0.8	0.5	0.5
Model 4	0.9	0.2	0.4	0.2
Model 5	0.9	0.1	0.1	0.1
Model 6	0.9	0.4	0.5	0.4

Table 4.2: Decay factor values for each model's network

The generator learning rate decay graph across four different model versions throughout the 20 training epochs can be observed in figure 4.2.

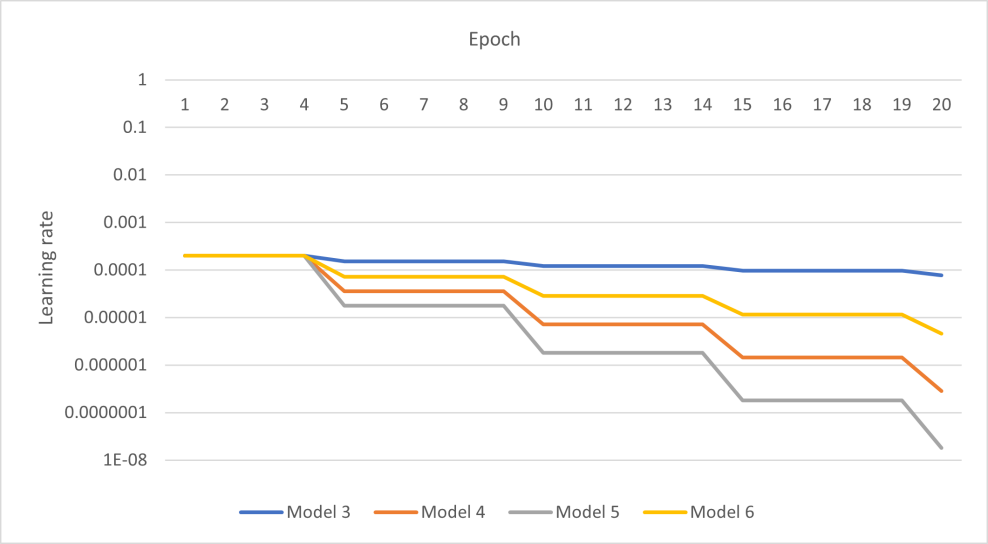

Figure 4.2: Generator learning rates decay across 4 different model versions

Similarly, the discriminator's learning rate graph relative to the 4 models throughout the 20 epochs is portrayed in figure 4.3.

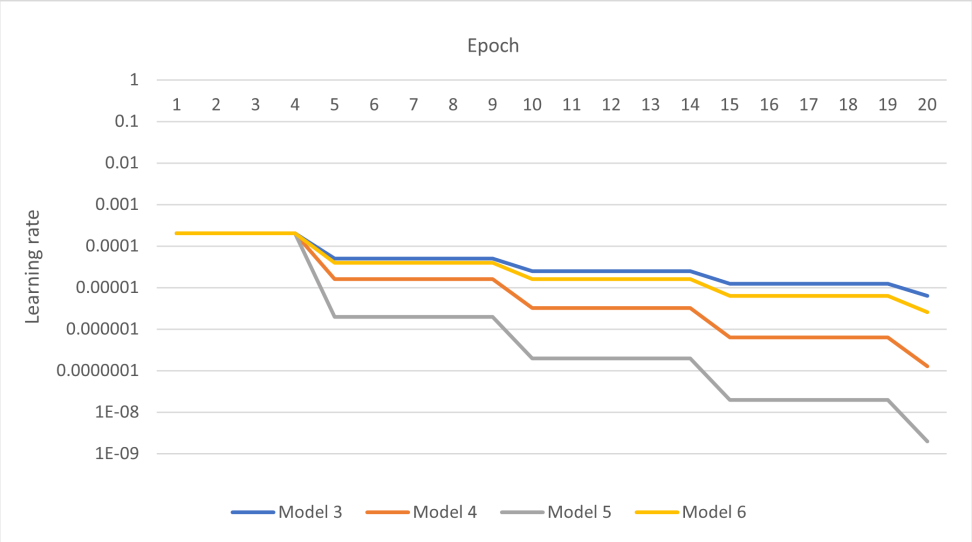

Figure 4.3: Discriminator learning rates decay across 4 different model versions

The FID values were then calculated in the full testing set, meaning that both low and medium quality images were enhanced.

4.3 Results and Discussion

This chapter aims to present and discuss the results obtained by the enumerated experimental settings. The outcomes will be explored systematically, shedding light on the key outcomes and patterns observed during experimentation and analysis. By examining the influence of both the training data and the learning rate scheduler in model development, this chapter will delve into the implications and potential limitations of the findings.

4.3.1 Influence of training data

As previously stated, each model underwent image enhancement and FID calculation steps for two different scenarios. In the first scenario, only low quality images from the testing set were enhanced by the models, with a baseline value of 53.013. The FID values obtained for Model 1 across all 20 epochs are illustrated in figure 4.4. Through its analysis, it is concluded that the model achieved its best value of 43.720 in epoch 4, representing an improvement of the baseline value. The model progressed consistently until epoch 5, from which the FID scores escalated abruptly, remaining this way until the last epoch.

Following the previous step, Model 2 was also firstly tested on image enhancement of only low quality images. The FID scores obtained can be observed in figure 4.4, in comparison with the baseline FID value and Model 1's results for the low quality testing set. The lowest value achieved by Model 2 was 39.771 in epoch 5, representing an improvement from the results achieved by Model 1. Similarly to that same model, the FID values shot up from epoch 6 onwards, suggesting that the model had already reached its optimal state.

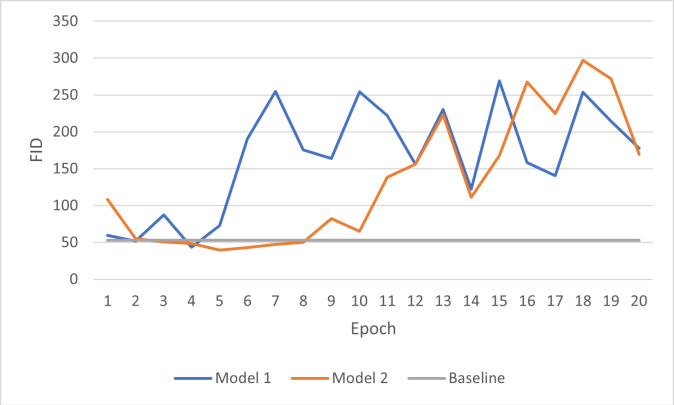

Figure 4.4: FID score obtained by models 1 and 2 tested on low quality images from the dataset

Regarding the second testing scenario, both models underwent evaluation on the generated enhanced images from the full testing set, containing both medium and low quality images. The achieved results from both models can be observed in figure 4.5 along with the baseline value of 27.440. Once again, Model 1 achieved its best result of 24.452 in epoch 4, this being the only value going under the baseline. On the other hand, Model 2 attained its lowest value of 22.484 in epoch 5, as expected.

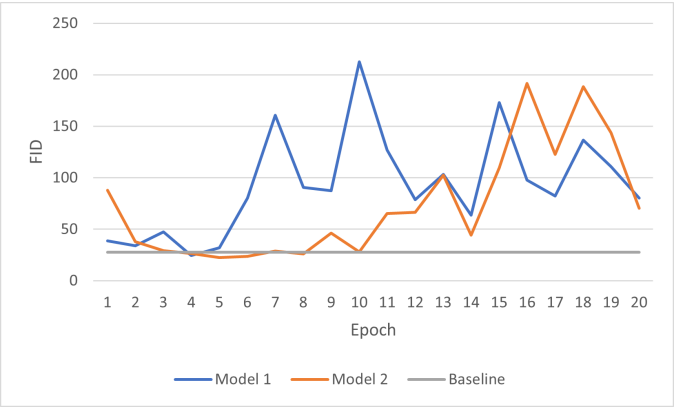

Figure 4.5: FID score obtained by Model 2 tested on low and medium quality images from the dataset

Through the analysis of the FID scores for both models in both situations, it is concluded that Model 2 attained more favourable results compared to Model 1, as depicted in table 4.3. While both models improved the baseline FID scores, implying that the image enhancement was successful in both situations, Model 2 achieved the lowest FID value from both models. Moreover, Model 2's scores remained under the baseline value for more epochs than Model 1 and showed more consistency in comparison, as observed in figures 4.4 and 4.5.

Furthermore, both models visually enhanced the original images, as represented in figure 4.6, where a real low quality image from the testing set is compared to its improved versions by each model. For this example, both images were generated using their respective models in their optimal

	Baseline	**Model 1**	**Model 2**
Lowest FID Score on Low Quality Testing Set	53.013	43.720	**39.771**
Lowest FID score on Low and Medium Quality Testing Set	27.440	24.452	**22.484**
Number of Images used During Training	-	10,905	17,272

Table 4.3: Models 1 and 2 comparison

state, this being epoch 4 for Model 1 and epoch 5 for Model 2. On one hand, the enhanced images presented a notable increase in high-frequency noise, with a higher increase by Model 1, which could lead to the potential presence of undesired artifacts. However, both demonstrated enhanced contrast in the walls of the left ventricle, increasing the differentiation between the ventricular wall and surrounding structures. This heightened contrast enables a clearer delineation of the left ventricle, which can be advantageous for accurate segmentation and analysis tasks. Therefore, generated images from both models showed improved clarity and sharpness overall in comparison to the original low quality images, increasing quality and therefore underlining the potential and efficiency of the presented methodology.

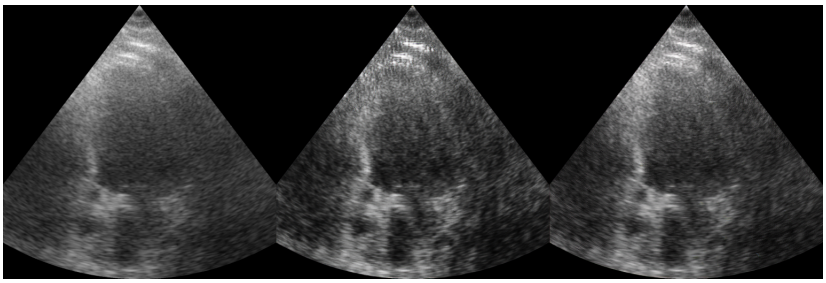

Figure 4.6: Visual comparison of a sample low quality original image and its enhancement by models 1 and 2 in their optimal state, respectively from left to right

Regarding the evolution of both models throughout the different epochs, in both situations, there was a consistent and favourable evolution until the optimal solutions in epochs 4 and 5, from which the values shot up, which could be explained by the discrepancy in loss values between the generator and discriminator, depicted in figure 4.7. A low discriminator loss indicates that the discriminator has become highly accurate in distinguishing between real and fake images, which, in turn, is detrimental to the generator's ability to further learn and improve, resulting in an imbalance during the training process.

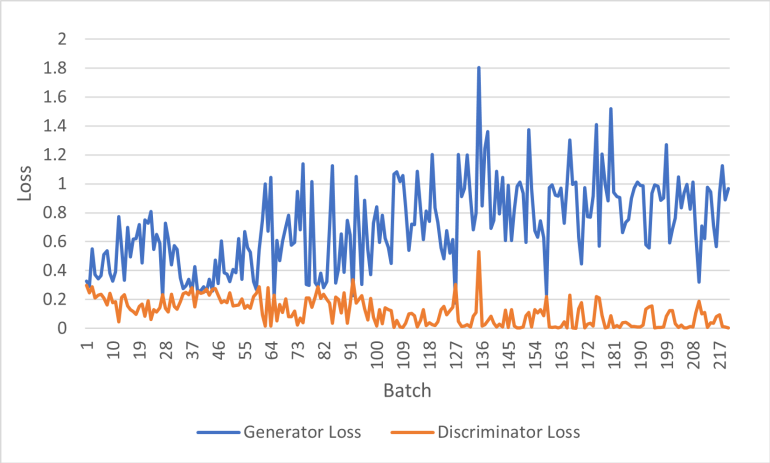

Figure 4.7: Model 2's generator and discriminator losses throughout training

The outcomes presented achieved positive results in comparison to the original images, highlighting the potential of CycleGAN in echocardiography image quality enhancement. The impact of the size and characteristics of the training set was studied and concluded that the usage of medium quality images during training was essential for model improvement. This enhances not only the model's adaptability to new data but also increases its overall precision and efficiency.

However, as previously stated, beyond epoch 5 until epoch 20, the values increase and become inconsistent, leading to a deterioration in the model and consequent quality of the generated images. Therefore, by incorporating techniques such as learning rate tuning, the model can be further improved, ensuring sustained enhancements to the quality and fidelity of the generated images throughout the training. This will allow the model to consistently and continuously converge towards the desired outcome, thereby maximizing the model's potential and attaining better results.

4.3.2 Influence of learning rate scheduler

Upon training and applying the obtained models in image generation, the FID scores across the 4 models were calculated and their evolution is depicted in figure 4.8, in comparison to the baseline. The minimum values obtained by each model are further specified in table 4.4, along with the corresponding epoch in which the value was attained.

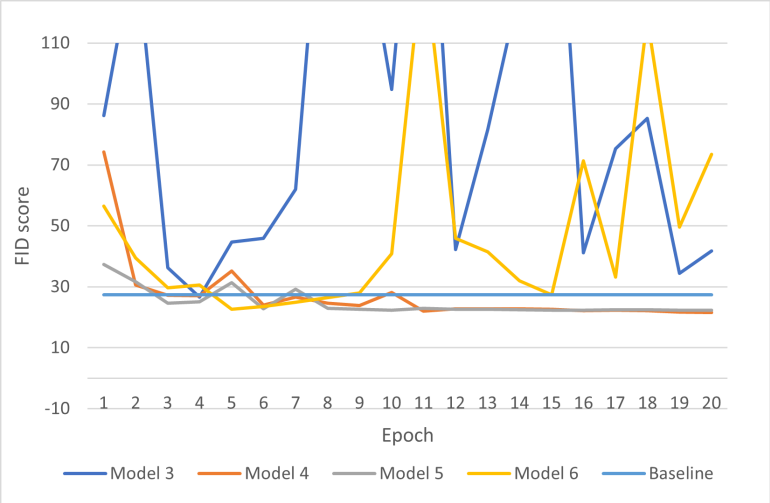

Figure 4.8: FID scores obtained by models 3 to 6 across all 20 epochs compared to the baseline

Through the analysis of the FID values across the 4 models, it can be concluded that Model 3, which presented a lower decay rate in both the generator and discriminator, obtained higher and more inconsistent values, suggesting that the model was not converging as expected. Similarly, model 6, which also evolved with a relatively low decay rate, attained slightly better results, however still inconsistent and higher than the baseline in most of its epochs. This improvement suggested that a higher decay rate could improve the FID score, which was then observed in Models 4 and 5.

Both models 4 and 5 succeeded in attaining convergence and remaining below the baseline value. Model 4 achieved better results of all 4 models, with a minimum score of 21.539 in contrast with its corresponding baseline of 27.440. This value was obtained in its 20th epoch, indicating that the model was continuously progressing and converging towards lower values and therefore attaining its objective.

	FID	**Epoch**
Model 3	26.660	4
Model 4	**21.539**	20
Model 5	22.332	15
Model 6	22.602	5

Table 4.4: Minimum FID scores attained by each model and respective epochs

The visual enhancement of the original images by Model 4 is further illustrated in figure 4.9. As previously discussed, the generated images present a higher noise ratio, which could be a potential negative aspect. Nevertheless, the images acquired higher contrast which allows for a better delineation of the heart structure and chamber delimitation. This improvement is valuable to achieve more accurate results in segmentation and diagnosis.

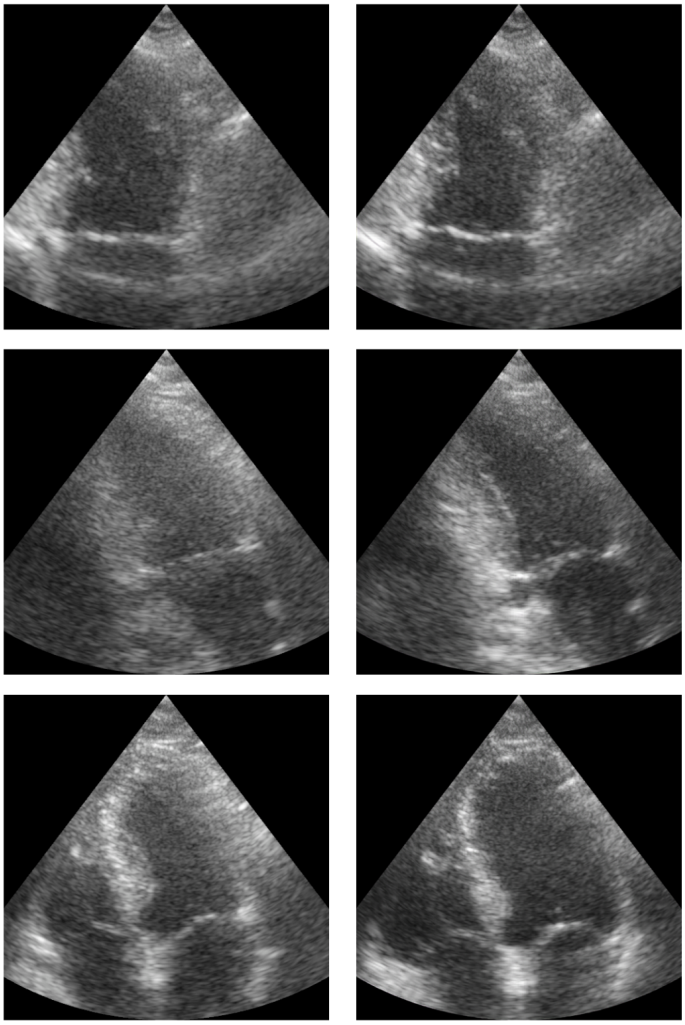

Figure 4.9: Comparison between the original images, on the left, and their respective enhanced versions by Model 4, on the right

The results obtained indicated that the incorporation and tuning of a learning rate scheduler was crucial in improving the model's overall convergence and performance. During the experimental process, the model achieved significant improvements regarding its FID score throughout the different epochs, indicating the successful enhancement of the generated images. The FID score reduction demonstrates that the model effectively learned the underlying features and structures of the input data, highlighting the increased proximity of lower quality images to higher quality ones.

The improved performance of the model can be attributed to the learning rate scheduler's capacity of dynamically adjusting the learning rate of each network independently throughout the training phase. By gradually reducing the learning rate throughout the epochs, the model could effectively achieve optimization whilst avoiding overshooting or becoming trapped in suboptimal solutions.

Despite attaining favorable and improved results, future research is possible by further tuning different learning rate schedulers or exploring the combination of multiple optimization techniques in order to improve the performance of CycleGAN for echocardiography image enhancement.

4.4 Conclusions

This chapter aims to conclude the cycleGAN development for echocardiographic image enhancement. In order to improve the model's performance, several modifications and adaptations were explored, ranging from a rigorous organization and preprocessing of the dataset to the application and tuning of an adequate learning rate scheduler.

The model architecture, loss functions, and hyperparameters were modified and adapted throughout the training phase. The incorporation of medium quality images during training and therefore increasing training data has attained positive results and shown to be crucial in model training. In order to improve convergence and avoid becoming trapped in local optima, a learning rate scheduler was also applied to gradually and dynamically adjust the learning rate of each network during training. Carefully altering the learning rate scheduler, as well as applying other modifications in the architectural context, led to significant increases in the model's performance.

The quantitative evaluation was conducted using the FID score. This method allowed for a quantitative assessment of the enhanced images, using original high quality testing images as reference. With an initial baseline score of 27.440 using the full testing set and a lowest value achieved during experimentation of 21.539, the generated images became visually and quantitatively closer to real high quality images, emphasising the effectiveness of the proposed methods.

While promising results have been attained, future research and improvement is possible and crucial for its expansion. The exploration of alternative model architectures and loss functions could potentially achieve more positive outcomes. Furthermore, dataset changes could enhance the model's capabilities to work with a wider range of echocardiographic images, such as incorporating additional training data from a wider and more diverse dataset.

The major purpose of this section of the study was to enhance the quality of echocardiographic images using a generative adversarial network, which led to visibly positive results, reinforcing the potential of cycleGAN in improving medical images. The accuracy and fidelity of the enhanced pictures has significantly improved as a consequence of the adjustments and adaptations performed. These outcomes lay the foundation for future investigation and development of image enhancing tools for cardiac diagnosis. Therefore, with continued efforts and expansion of this field, it is anticipated that automated echocardiographic image enhancement techniques will be crucial in enhancing patient care and results in the cardiovascular health sector.

Chapter 5

Echocardiographic Quality Enhancement in Automatic Segmentation

Automatic segmentation algorithms employ various advanced computational techniques aiming to automatically delineate specific structures. In automatic echocardiography segmentation, the objective consists on accurately and precisely delineating the boundaries of cardiac structures, contributing to an improvement on medical analysis and diagnosis.

Through the utilization of sophisticated image processing and machine learning methodologies, these algorithms analyze the pixel-level information present in the echocardiographic images. They aim to identify and differentiate the specific anatomical regions of interest within the cardiac structures, thereby allowing an accurate delineation and localization of these regions.

These tasks are generally performed by physicians, which can be time consuming, highlighting the need for complementary or full automatic analysis tools. Despite the attempts to enhance these automatic methods, the image quality remains a significant limitation that diminishes their success rate. In this regard, the application of image enhancement techniques emerges as a valuable approach to address this challenge.

This chapter encompasses the application of the models derived from this research to automatic segmentation algorithms, aiming to enhance the quality of the segmentation task. Its focus will be the integration of enhanced echocardiography sequences, improved using the model developed in section 3, into the EchoNet-Dynamic [29] framework, a video-based automatic segmentation algorithm, aiming to evaluate the impact of enhanced data on the accuracy of left ventricle segmentation.

The methodology and integration process will be thoroughly analysed along its results, which will be compared to the performance of automatic segmentation in the original sequences. This analysis aims to assess the impact of image quality enhancement in improving the accuracy and reliability of cardiac function assessment using automatic segmentation such as EchoNet-Dynamic framework.

5.1 Methodology

The present section aims to delve into the methodology employed along with its limitations. The resources and experimental settings will be extensively explored, providing a comprehensive overview of the methodology adopted, highlighting its strengths, limitations, and the context in which the study was conducted.

5.1.1 Automatic LV Segmentation

The field of cardiac function assessment has traditionally relied on human experts, whose assessments are limited in scope and subject to inter-observer variability. In order to address these limitations, several studies have been conducted, from which EchoNet-Dynamic [29], a video-based deep learning algorithm, can be highlighted as an effective approach example. The author's contribution therefore lies in validating and exploring the integration of enhanced images within the existing EchoNet-Dynamic framework.

Developed by Ouyang et. al, EchoNet-Dynamic aims to automate critical tasks such as left ventricle segmentation, ejection fraction estimation and cardiomyopathy assessment. EchoNet-Dynamic achieved high accuracy in left ventricle segmentation with a Dice similarity coefficient of 0.92. Moreover, it attained a mean absolute error of 4.1% regarding ejection fraction prediction while also reliably classifying heart failure with reduced ejection fraction, reaching an area under the curve of 0.97. The algorithm's performance remains effective when applied to an external dataset, with a mean absolute error of 6.0% in ejection fraction prediction and an area under the curve of 0.96 in classifying heart failure with reduced ejection fraction.

Dataset Analysis and Specifications

The dataset used in the development of EchoNet-Dynamic [29] comprises 10,030 echocardiography videos obtained from 10,030 unique individuals between 2016 and 2018 at Stanford Health Care. Each video corresponds to an apical four-chamber view extracted from a full resting echocardiogram study, capturing the heart from different angles and using various image acquisition techniques. The videos were randomly divided into training (7,465 videos), validation (1,277 videos), and testing (1,288 videos) sets.

In order to safeguard patient privacy and remove any identifiable information, a meticulous data curation process was implemented. Therefore, an automated preprocessing workflow was employed, performing tasks such as the cropping and masking of videos to eliminate elements such as text, electrocardiogram readings, respirometer information, and data beyond the scanning sector.

Model Architecture

EchoNet-Dynamic [29] was developed and trained using Python and the PyTorch library. The model employed the Deeplabv3 architecture for semantic segmentation, with a base architecture

of a 50-layer residual net. It was trained using stochastic gradient descent optimizer and minimized pixel-level binary crossentropy loss.

In order to incorporate spatiotemporal convolutions, Ouyang et. al tested three different model architectures: R3D, MC3, and R2+1D. Ultimately, the R2+1D architecture, which decomposes the spatiotemporal convolutions, was selected for EchoNet-Dynamic due to its stronger performance. The R3D architecture considered spatial and temporal dimensions jointly, while the MC3 and R2+1D architectures found a middle ground between two-dimensional and full three-dimensional convolutions.

For ejection fraction prediction, the authors aimed to train the models as a way to minimize the squared loss between the predicted and true ejection fraction values. Stochastic gradient descent optimizer with an initial learning rate of 0.0001, momentum of 0.9, and a batch size of 16 was used for training, with the learning rate decaying by a factor of 0.1 every 15 epochs.

During training, video clips of 32 frames were generated by sampling certain frames with a sampling period of 2. Data augmentation techniques, such as padding each training video clip and taking random crops of the original frame size, were employed by the authors so as to increase the dataset size while also introducing diversity. The model's architecture is schematized in figure 5.1.

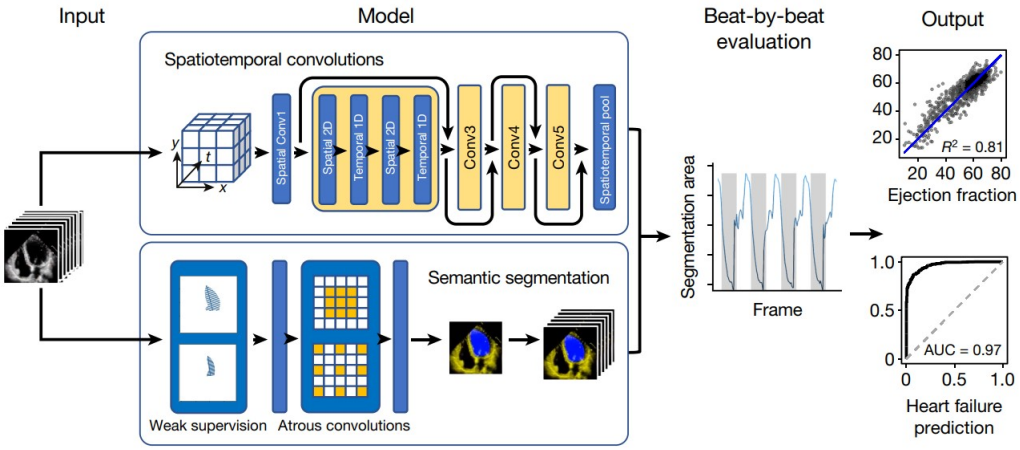

Figure 5.1: Echonet model architecture [29]

5.1.2 Evaluation Metrics

In order to numerically assess the efficiency of the enhanced images in automatic cardiac segmentation, quantitative evaluation metrics such as the Dice similarity coefficient (DICE), Mean absolute deviation (MAD) and Hausdorff Distance (HD) were applied.

The DICE coefficient aims to quantify the overlapping of two sets of binary masks, in this context, the segmented regions obtained from the enhanced images and the real annotations, as illustrated in figure 5.2. The DICE value ranges from 0 to 1, where a score of 0 indicates no overlapping between both sets, while a score of 1 represents a perfect overlap. Therefore, the higher the score, the closer the results from the generated images align with the ground truth annotations,

indicating a higher accuracy. Thus, this evaluation metric allows for a quantitative assessment of the enhancement performed by the cycleGAN models while also assessing its capability of preserving the original cardiac structure, which remains a key point given the specificity of the data.

$$\text{Dice} = \frac{2\ X\ Area\ of\ overlap}{Total\ area} = \frac{2 \times}{}$$

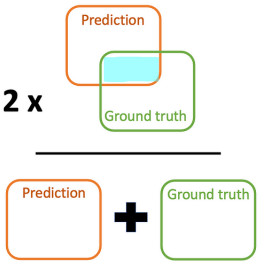

Figure 5.2: Dice Coefficient diagram [14]

To further assess the model's performance variation regarding original and enhanced data, two additional evaluation metrics were employed, these being the Hausdorff Distance (HD) and Mean Absolute Deviation (MAD).

The HD aims to measure the dissimilarity between two sets of points, and therefore quantifying the maximum distance between corresponding points between the generated segmentation and its ground truth, as schematized in figure 5.3. Therefore, a lower HD score indicates a shorter distance and consequently a better overlap with the ground truth.

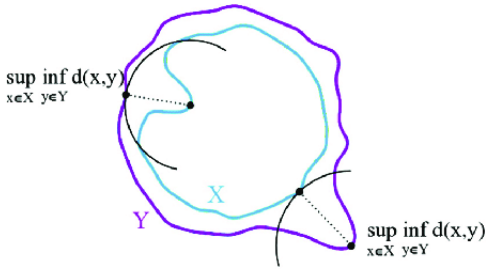

Figure 5.3: HD diagram [34]

Similarly, MAD measures the average absolute difference between the corresponding points in the generated segmentation and its ground truth. It provides a quantitative assessment of the general accuracy of the segmentation, where a lower MAD score indicates a closer match to the ground truth. It is calculated as follows, where n represents the number of data points, x_i refers to each individual data point, and $\bar{x}$ represents the mean of the data set.

$$MAD = \frac{1}{n}\sum_{i=1}^{n}|x_i - \bar{x}|$$

5.2 Experiments

The resulting data from the application of cycleGAN in echocardiography image enhancement was integrated in the Echonet-dynamic model. It aimed to test its efficiency in automatic segmentation tools, thus seeking further evaluation and testing its performance in a practical context.

Therefore, automatic left ventricular segmentation was tested on low and medium quality sequence videos from the CAMUS [23] dataset, comprising the end diastolic and end systolic frames for each sequence, in order to understand the efficiency of the developed methods. Automatic segmentation was initially applied to the unmodified sequences, followed by their counterparts enhanced by model 3, which had proven to be the most efficient among the developed models in the previous section. A total of 50 low or medium quality sequences from the testing set were used. The segmentation model, Echonet-dynamic, used pretrained weights to automatically generate the left ventricular segmentations for each sequence frame. The annotations from the CAMUS dataset were utilized in order to establish the groundtruth for each sequence, and therefore evaluate the results' accuracy.

5.3 Results and Discussion

Following the application of both original and enhanced sequences in Echonet-dynamic, an automatic segmentation was predicted for each image, as illustrated in figure 5.4.

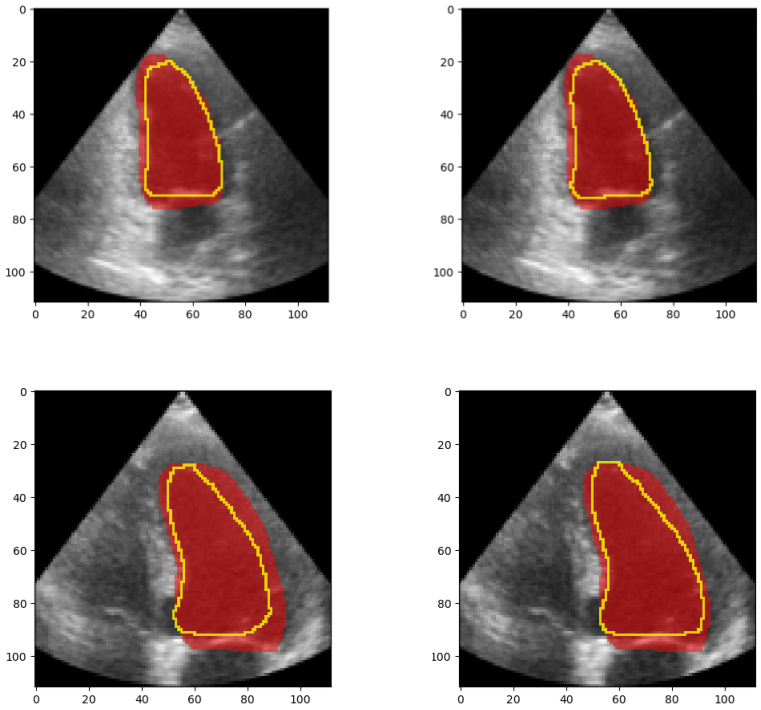

Figure 5.4: Automatic segmentation performed by Echonet-Dynamic in original images (left) and their respective enhanced versions (right). The red mask represents the groundtruth and the yellow contour the predicted segmentation

The segmentation results were then quantitatively evaluated by calculating the DICE, MAD and HD scores, in order to evaluate its efficiency and improvement.

The DICE values obtained from the generated images compared to the original data are portrayed in table 5.1 and its distribution is pictured in figure 5.5. The overall DICE values calculated by the model in different stages are indicated, comprising both the end of systole and the end of diastole. Additionally, both the highest and lowest values attained are highlighted in the table, effectively delineating the range within which the results were observed. Through its analysis, it is observed that the DICE score of the generated images consistently exhibited higher values across the different categories, attaining an overall improvement of approximately 14.06% relatively to the original images.

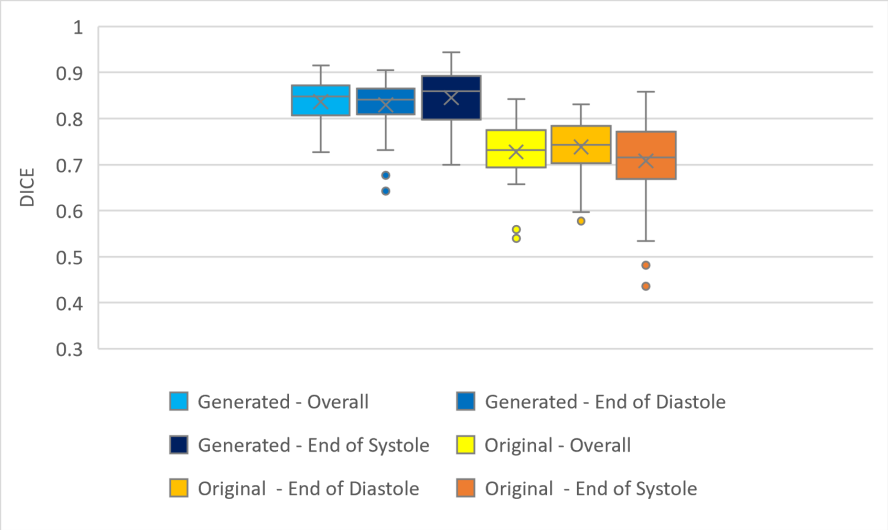

Figure 5.5: DICE distribution across patient data

	Original	**Generated**
Overall Dice Value	0.7335	0.8367
Overall End of Diastole Dice Value	0.7426	0.8291
Overall End of Systole Dice Value	0.7204	0.8469
Highest Dice Value Attained	0.8417	0.9439
Lowest Dice Value Attained	0.5399	0.6998

Table 5.1: Dice similarity coefficient comparison between original and generated images

Furthermore, MAD and HD scores were also calculated for the same samples. The average values achieved for MAD are indicated in table 5.2 and its distribution is depicted in figure 5.6.

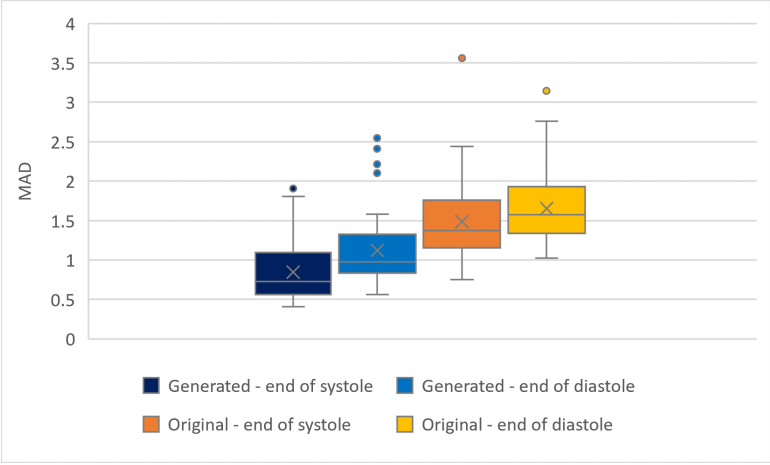

Figure 5.6: MAD scores distribution across patient data

Similarly, the HD average values are described in table 5.3 and its respective distribution is illustrated in figure 5.7.

	Original	**Generated**
Average End of Systole MAD	1.486	0.846
Average End of Diastole MAD	1.652	1.121

Table 5.2: Average MAD values comparison between original and generated images

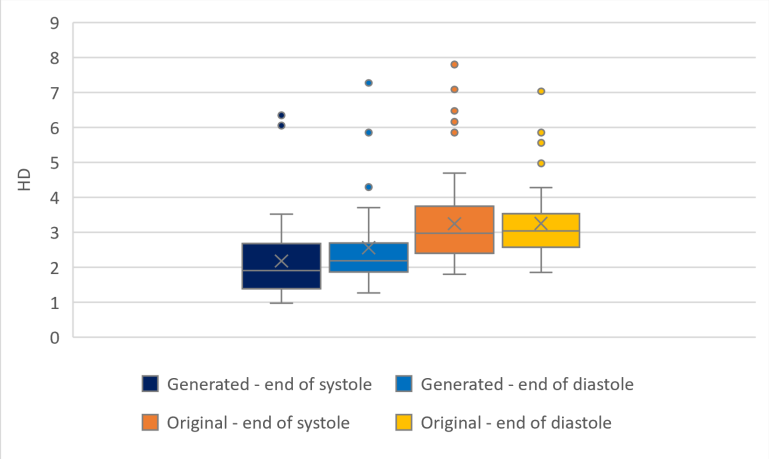

Figure 5.7: HD scores distribution across patient data

	Original	**Generated**
Average End of Systole HD	3.242	2.176
Average End of Diastole HD	3.249	2.557

Table 5.3: Average HD values comparison between original and generated images

Similarly to the DICE results, the sequences enhanced by Model 3 attained better results for both MAD and HD metrics across both systole and diastole stages. Through the usage of quality improved sequences, the MAD value suffered an average reduction of 43.02% regarding the end of systole and 31.99% for the end of diastole. Similarly, the HD metric achieved a reduction of 32.82% for the end of systole and 21.29% for end of diastole. This indicates a stronger similarity of the predicted segmentation to the ground truth, highlighting the ability of the model to automatically determining the cardiac anatomy when employing enhanced images. These results underline the potential of cycleGAN-based image enhancement techniques in the application of automatic segmentation tools, aiding its efficiency and contributing to more accurate results.

5.4 Conclusions

This chapter aimed to explore the application of a cycleGAN-based image enhancement model to improve the quality of echocardiography images for automatic segmentation purposes. The original images contain inherent limitations which can affect the performance of segmentation

algorithms, remaining a major drawback regarding their efficiency. By applying the proposed image enhancement model, the quality of the input images was significantly improved.

The results obtained from the application of the enhanced sequences to Echonet-dynamic demonstrated the effectiveness of the image improvement approach. The enhanced images led to increased segmentation results, closer to the ground truth values, in comparison to the original ones, addressing the challenge of image quality as a major obstacle in automatic segmentation algorithms.

The outcomes of this study highlight the potential of using CycleGAN-based image enhancement techniques to enhance echocardiography images and improve the performance of subsequent segmentation algorithms. The advancements achieved positively contribute to the field of cardiac image analysis, while the improved image quality obtained through this technique holds important implications for the accurate and reliable assessment of cardiac function.

Despite the positive results attained, future research is crucial to the continuous development of the field. Further improvements could be achieved by refining the image enhancement model, exploring different architectures and investigating the impact of enhanced images on other cardiac imaging tasks. Overall, this research and application demonstrates the promising role of image enhancement in overcoming image quality limitations in automatic segmentation techniques and advancing the field of echocardiography analysis.

Chapter 6

Conclusion and Future Work

6.1 Conclusions

This dissertation had as its primary objective the improvement of the quality and interpretability of echocardiography images, which hold major importance in the diagnosis and monitoring of various cardiac conditions. In order to achieve this, a cycleGAN approach was employed. Through extensive experimentation, variables such as different training data and learning rate schedulers, were tested and thoroughly evaluated, as long as their impact in the model's performance. The findings have indicated a successful enhancement of the images, with a clear visual improvement, showing improved clarity and sharpness overall, and an FID decrease of 21.48% compared to the baseline. Moreover, the integration of the enhanced images into an automatic segmentation algorithm was conducted, measuring the viability of the improved images in a practical setting. This experiment led to positive results, improving the accuracy and efficiency in automatic cardiac structure delineation. Overall, this research contributes substantially to the field of echocardiography image analysis by presenting a comprehensive framework for image enhancement and its application in automatic segmentation. The developed CycleGAN model has the capacity to increase the accuracy and efficiency of cardiac analysis, thereby benefiting patients and healthcare professionals in the diagnosis of cardiovascular pathologies.

6.2 Future Work

This research therefore represents a significant stride in addressing the challenges associated with echocardiography image quality and automatic segmentation. However, further investigations and improvements remain viable avenues for future exploration. Subsequent studies could delve into additional variables, such as diverse network architectures or advanced loss functions, in order to further enhance the performance of the CycleGAN model. A gathering of different datasets could also play a crucial role in model training, allowing the model to broaden its adaptation abilities and acquire a broader spectrum of information, improving its generalization capabilities. This would allow for a more comprehensive understanding of the underlying patterns within the existing data

and therefore enhancing its data handling capacity. Additionally, a further exploration and tuning of the learning rate scheduler could potentially lead to improved results, by enhancing convergence and achieving a better balance overall. The iterative process of experimentation and optimization is crucial for maximizing the performance of the model and achieving better results in future applications.

References

[1] American Heart Association. Single photon emission computed tomography (spect), Jul 2015.

[2] Mass General Brigham. Echocardiography: Cooley dickinson radiology amp; imaging, 2023.

[3] Antonia Creswell, Tom White, Vincent Dumoulin, Kai Arulkumaran, Biswa Sengupta, and Anil A. Bharath. Generative adversarial networks: An overview. *IEEE Signal Processing Magazine*, 35(1):53–65, 2018.

[4] CycleGAN. Cyclegan, 2022.

[5] MAURIZIO DE ANGELIS. Heart anatomy by maurizio de angelis/science photo library, Nov 2019.

[6] Gerhard-Paul Diller, Astrid Lammers, Sonya Babu-Narayan, Wei Li, Robert M. Radke, Helmut Baumgartner, Michael Gatzoulis, and Stefan Orwat. Denoising and artefact removal for transthoracic echocardiographic imaging in congenital heart disease: utility of diagnosis specific deep learning algorithms. *The International Journal of Cardiovascular Imaging*, 35, 12 2019.

[7] DocPanel. How any imaging center can offer cardiac mri, 2020.

[8] Echelon. Heart ct scan in london: Ct scanner, Jan 2023.

[9] Maria Escobar, Angela Castillo, Andrés Romero, and Pablo Arbeláez. Ultragan: Ultrasound enhancement through adversarial generation. In Ninon Burgos, David Svoboda, Jelmer M. Wolterink, and Can Zhao, editors, *Simulation and Synthesis in Medical Imaging*, pages 120–130, Cham, 2020. Springer International Publishing.

[10] Manuel Franco, Richard S. Cooper, Usama Bilal, and Valentín Fuster. Challenges and opportunities for cardiovascular disease prevention. *The American Journal of Medicine*, 124(2):95–102, 2011.

[11] Roy Ganz. A review of generative adversarial networks-part 1, Dec 2020.

[12] Donald Geman, Stuart Geman, Neil Hallonquist, and Laurent Younes. Visual turing test for computer vision systems. *Proceedings of the National Academy of Sciences of the United States of America*, 112, 03 2015.

[13] Martin Heusel, Hannes Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Advances in Neural Information Processing Systems*, pages 6626–6637, 2017.

[14] Nghi Huynh. Understanding evaluation metrics in medical image segmentation, Mar 2023.

[15] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks, 2016.

[16] Mohammad Jafari, Hany Younan Azer Girgis, Nathan Woudenberg, Nathaniel Moulson, Christina Luong, Andrea Fung, Shane Balthazaar, John Jue, Micheal Tsang, Justin Bribe, Kenneth Gin, Robert Rohling, Purang Abolmaesumi, and Teresa Tsang. Cardiac point-of-care to cart-based ultrasound translation using constrained cyclegan. *International Journal of Computer Assisted Radiology and Surgery*, 15, 04 2020.

[17] Mohammad H. Jafari, Zhibin Liao, Hany Girgis, Mehran Pesteie, Robert Rohling, Ken Gin, Terasa Tsang, and Purang Abolmaesumi. Echocardiography segmentation by quality translation using anatomically constrained cyclegan. In Dinggang Shen, Tianming Liu, Terry M. Peters, Lawrence H. Staib, Caroline Essert, Sean Zhou, Pew-Thian Yap, and Ali Khan, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, pages 655–663, Cham, 2019. Springer International Publishing.

[18] Shizuo Kaji and Satoshi Kida. Overview of image-to-image translation using deep neural networks: denoising, super-resolution, modality-conversion, and reconstruction in medical imaging, 05 2019.

[19] Mehmet Kocak. Computed tomography, Jan 2023.

[20] Mehmet Kocak. Magnetic resonance imaging, Jan 2023.

[21] Mehmet Kocak. Positron emission tomography (pet), Jan 2023.

[22] Huong T Le, Nicholas Hangiandreou, Robert Timmerman, Mark J Rice, W Brit Smith, Lori Deitte, and Gregory M Janelle. Imaging artifacts in echocardiography. *Anesthesia and analgesia*, 122(3):633—646, March 2016.

[23] Sarah Leclerc, Erik Smistad, João Pedrosa, Andreas Østvik, Frederic Cervenansky, Florian Espinosa, Torvald Espeland, Erik Andreas Rye Berg, Pierre-Marc Jodoin, Thomas Grenier, Carole Lartizien, Jan D'hooge, Lasse Lovstakken, and Olivier Bernard. Deep learning for segmentation using an open large-scale dataset in 2d echocardiography. *IEEE Transactions on Medical Imaging*, 38(9):2198–2210, 2019.

[24] Zhibin Liao, Mohammad Jafari, Hany Younan Azer Girgis, Kenneth Gin, Robert Rohling, Purang Abolmaesumi, and Teresa Tsang. *Echocardiography View Classification Using Quality Transfer Star Generative Adversarial Networks*, pages 687–695. 10 2019.

[25] Jackson JK. Mc Namara K, Alzubaidi H. Cardiovascular disease as a leading cause of death: how are pharmacists getting involved? vol. 8(p. 1-11), February 2019.

[26] MedScience. Spect imaging applications to myocardial perfusion and brain imaging, Sep 2021.

[27] Natalia Natalinova, D.K. Avdeeva, Veniamin Kazakov, Vladimir Baranov, Olga Galtseva, and Denis Ivashkov. Computer spatially oriented reconstruction of a 3d heart shape based on its tomographic imaging. *MATEC Web of Conferences*, 79:01005, 01 2016.

[28] Catherine M. Otto. *Textbook of clinical echocardiography*. Elsevier, 2018.

[29] David Ouyang, Bryan He, Amirata Ghorbani, Neal Yuan, Joseph E Ebinger, Curtis P Langlotz, Paul A Heidenreich, Robert A Harrington, David Liang, Euan A Ashley, and James Zou. Video-based ai for beat-to-beat assessment of cardiac function. 580(7802):252–256, Mar 2020.

[30] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.

[31] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training gans. *arXiv preprint arXiv:1606.03498*, 2016.

[32] M. J. Schuuring, I. Išgum, B. Cosyns, S. A. J. Chamuleau, and B. J. Bouma. Routine echocardiography and artificial intelligence solutions. vol. 8(p. 648877), February 2021.

[33] Michael J. Shea Thomas Cascino. Echocardiography. 2021.

[34] Francesca Uccheddu, Michaela Servi, Rocco Furferi, and Lapo Governi. Comparison of mesh simplification tools in a 3d watermarking framework. pages 60–69, 05 2018.

[35] UOHI. Pet myocardial perfusion imaging, 2023.

[36] WHO. Cardiovascular diseases, 2023.

[37] Xin Yi, Ekta Walia, and Paul Babyn. Generative adversarial network in medical imaging: A review. *Medical Image Analysis*, 58:101552, 2019.

[38] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2242–2251, 2017.