Improving Image Quality in Cardiac Computed Tomography using Deep Learning

David Wajngot

Supervisors: Vikas Gupta, Mariana Bustamante
Examiner: Tino Ebbers
Upphovsrätt


Copyright

The publishers will keep this document online on the Internet – or its possible replacement – for a period of 25 years starting from the date of publication barring exceptional circumstances. The online availability of the document implies permanent permission for anyone to read, to download, or to print out single copies for his/hers own use and to use it unchanged for non-commercial research and educational purpose. Subsequent transfers of copyright cannot revoke this permission. All other uses of the document are conditional upon the consent of the copyright owner. The publisher has taken technical and administrative measures to assure authenticity, security and accessibility. According to intellectual property law the author has the right to be mentioned when his/her work is accessed as described above and to be protected against infringement. For additional information about the Linköping University Electronic Press and its procedures for publication and for assurance of document integrity, please refer to its www home page [http://www.ep.liu.se/](http://www.ep.liu.se/).

© David Wajngot
Abstract

Cardiovascular diseases are the largest mortality factor globally, and early diagnosis is essential for a proper medical response. Cardiac computed tomography can be used to acquire images for their diagnosis, but without radiation dose reduction the radiation emitted to the patient becomes a significant risk factor. By reducing the dose, the image quality is often compromised, and determining a diagnosis becomes difficult. This project proposes image quality enhancement with deep learning. A cycle-consistent generative adversarial neural network was fed low- and high-quality images with the purpose to learn to translate between them. By using a cycle-consistency cost it was possible to train the network without paired data. With this method, a low-quality image acquired from a computed tomography scan with dose reduction could be enhanced in post processing.

The results were mixed but showed an increase of ventricular contrast and artifact mitigation. The technique comes with several problems that are yet to be solved, such as structure alterations, but it shows promise for continued development.
Acknowledgments

My gratitude goes out to the entire team at KVM for a great semester. To my supervisor, Vikas Gupta, for constantly sharing his knowledge and kindness. To my other supervisor, Mariana Bustamante, for all the great advice, and for answering hundreds of questions about everything and anything. To Tino Ebbers, for being the most sympathetic examiner a master student could ask for. To my opponent, Camilla Almquist, for always lending me an ear.

David Wajngot
Linköping, 2019
Contents

Abstract iii
Acknowledgments iv
Contents v
List of Figures vii
List of Tables ix

1 Introduction 1
  1.1 Motivation ................................................. 1
  1.2 Aim .................................................. 2
  1.3 Background ............................................ 2
  1.4 Research Questions ..................................... 2
  1.5 Delimitations ......................................... 2

2 Theory and Related Work 3
  2.1 Computed Tomography .................................... 3
    2.1.1 The Linear Attenuation Coefficient ................. 4
    2.1.2 Artifacts ........................................ 5
  2.2 Image Quality Metrics ................................ 6
    2.2.1 Signal to Noise Ratio .............................. 6
    2.2.2 Contrast to Noise Ratio ........................... 7
  2.3 Deep Learning ........................................ 7
    2.3.1 Neural Networks ................................... 8
    2.3.2 Supervised vs Unsupervised Learning .............. 9
    2.3.3 The Cost Function ................................ 10
    2.3.4 Gradient Descent and Backpropagation ............ 11
  2.4 Convolutional Neural Networks .......................... 12
  2.5 Generative Adversarial Networks ...................... 12
    2.5.1 Cycle-Consistent Generative Adversarial Networks 13
  2.6 Histogram Equalization ................................ 16
    2.6.1 Adaptive Histogram Equalization ................ 17
  2.7 Related Work .......................................... 18

3 Method 20
  3.1 Image Acquisition .................................... 20
  3.2 Classification of Images ................................ 21
  3.3 CycleGAN ............................................ 21
  3.4 Evaluation of Results .................................. 22
    3.4.1 Image Quality Metrics ............................ 22
    3.4.2 Qualitative Test ................................ 23
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>How the CT acquires images by gantry rotation</td>
</tr>
<tr>
<td>2.2</td>
<td>A cardiac CT image with an ICD from the data set. The artifact is apparent and reduces image quality considerably</td>
</tr>
<tr>
<td>2.3</td>
<td>The neuron, the elementary unit of a neural network</td>
</tr>
<tr>
<td>2.4</td>
<td>Illustration of a simple example of the neural network. It usually consists of neurons in many layers</td>
</tr>
<tr>
<td>2.5</td>
<td>The structure of a conventional CNN</td>
</tr>
<tr>
<td>2.6</td>
<td>The structure of a conventional GAN</td>
</tr>
<tr>
<td>2.7</td>
<td>A pair of images from two different domains. An example of paired data [19]</td>
</tr>
<tr>
<td>2.8</td>
<td>A visualization of the CycleGAN</td>
</tr>
<tr>
<td>2.9</td>
<td>A visualization of how the cycle-consistency cost works in the CycleGAN</td>
</tr>
<tr>
<td>2.10</td>
<td>A gray scaled image and its corresponding histogram. The overall contrast is increased</td>
</tr>
<tr>
<td>2.11</td>
<td>A gray scaled image when it is processed with HE.</td>
</tr>
<tr>
<td>2.12</td>
<td>The image after HE and its corresponding histogram</td>
</tr>
<tr>
<td>2.13</td>
<td>The image after CLAHE and its corresponding histogram</td>
</tr>
<tr>
<td>4.1</td>
<td>The results from enhancement network 1. Epochs are presented on the x-axis and the SNR and CNR values are presented on the y-axis. Every 5th epoch is included from 1 to 80 (including 80)</td>
</tr>
<tr>
<td>4.2</td>
<td>Test images from enhancement network 1. Original images are displayed to the left and their corresponding processed versions are on the right.</td>
</tr>
<tr>
<td>4.3</td>
<td>Test images from enhancement network 1. Original images are displayed to the left and their corresponding processed versions are on the right.</td>
</tr>
<tr>
<td>4.4</td>
<td>The results from enhancement network 2. Epochs are presented on the x-axis and the SNR and CNR values are presented on the y-axis. Every 5th epoch is included from 1 to 80 (including 80)</td>
</tr>
<tr>
<td>4.5</td>
<td>Test images from enhancement network 2. Original images are displayed to the left and their corresponding processed versions are on the right.</td>
</tr>
<tr>
<td>4.6</td>
<td>Test images from enhancement network 2. Original images are displayed to the left and their corresponding processed versions are on the right.</td>
</tr>
<tr>
<td>4.7</td>
<td>The results from enhancement network 3. Epochs are presented on the x-axis and the SNR and CNR values are presented on the y-axis. Every 5th epoch is included from 1 to 80 (including 80)</td>
</tr>
<tr>
<td>4.8</td>
<td>Test images from enhancement network 3. Original images are displayed to the left and their corresponding processed versions are on the right.</td>
</tr>
<tr>
<td>4.9</td>
<td>Test images from enhancement network 3. Original images are displayed to the left and their corresponding processed versions are on the right.</td>
</tr>
<tr>
<td>4.10</td>
<td>The SNR percentage difference for the various methods</td>
</tr>
<tr>
<td>4.11</td>
<td>The CNR percentage difference for the various methods</td>
</tr>
<tr>
<td>4.12</td>
<td>Test images processed with HE. Original images are displayed to the left and their corresponding processed versions are on the right</td>
</tr>
</tbody>
</table>
4.13 Test images processed with CLAHE. Original images are displayed to the left and their corresponding processed versions are on the right. ................................................. 36
4.14 Test images from the artifact reduction network. Original images are displayed on the left and their corresponding processed versions are on the right. ......................... 37
4.15 Test images from the artifact reduction network. Original images are displayed on the left and their corresponding processed versions are on the right. ......................... 38
4.16 Histograms representing the results from test 2: contrast. The medical students are represented to the left and the more experienced radiologists to the right. ......................... 40
4.17 Histograms representing the results from test 2: noise. The medical students are represented to the left and the more experienced radiologists to the right. ......................... 40
4.18 Histograms representing the results from test 2: artifacts. The medical students are represented to the left and the more experienced radiologists to the right. ......................... 41
4.19 Histograms representing the results from test 2: contrast in artifact images. The medical students are represented to the left and the more experienced radiologists to the right. .... 41

5.1 The plan for the final pipeline of the finished product. This was not included in the scope of the project. .......................................................... 48
List of Tables

4.1 HE and CLAHE metrics .................................................................................. 34
4.2 Results from the inexperienced evaluators from qualitative test part 1. .......... 39
4.3 Results from the experienced evaluators from qualitative test part 1. .......... 39
4.4 Criteria grades .............................................................................................. 39
4.5 Results from qualitative test part 3. Visualizes how many processed images that were 
   diagnostically preferred of the total 40 by the medical students. ..................... 42
4.6 Results from qualitative test part 3. Visualizes how many processed images that were 
   diagnostically preferred of the total 40 by the experienced clinicians. .......... 42
Chapter 1

Introduction

Cardiovascular diseases are the largest mortality factor globally, claiming more victims every year than any other cause. People who suffer from a cardiovascular disease or are on the verge of becoming ill are in great need of early diagnosis, the earlier the better. With an early diagnosis the effect of most cardiovascular diseases can be mitigated and prevented by focusing on decreasing risk factors originating from unhealthy behaviour. Early detection enables proper medical management of the disease, such as counselling, medicine and continuous diagnosis during follow-ups [1].

Modern day computed tomography (CT) has a very high diagnostic accuracy. The patient typically only has to be in the CT scan for only a few minutes (depending on the purpose of the scan) and high quality images can still be acquired. The advantages of CT has for long been deemed worth the risks of the radiation to which the patient is exposed. However, the consensus within the medical community is to maintain the radiation dose as low as possible. The consequences of lowering the radiation dose however, is often the loss of image quality. There is still much ongoing research regarding methods to lower the radiation dose while maintaining the image quality high, and there is still a desire to improve CT because of its frequent usage within the medical community around the world [2] [3].

1.1 Motivation

Medical imaging has been an important part of diagnostics since the late 1940s when it was initialized with the conventional radiography, the X-ray. Since then it has been developed in many different directions, not only to overcome the many challenges of the area, but also to advance the knowledge and understanding of normal anatomy and physiology. In 1971 it was possible for the first time to non-invasively acquire images of internal anatomy within a relatively short span of time by using CT [2]. CT is now used widely in modern medicine. It utilizes X-ray radiation that rotates in an arc around the body to produce a three dimensional image of the area of interest. The patient is placed into a CT scanner where the information is collected [4].

While Magnetic Resonance Imaging (MRI) generally is the preferred technique when high soft tissue contrast is required, CT acquires images significantly faster. For this reason CT is often used for initial, faster inspection (such as in the emergency room) with a follow-up of MRI that is slower but better suited for more detailed diagnosis. MRI is the primary choice when it comes to getting better detail for imaging in the soft tissue, such as ligament and tendon injuries. However, CT is the preferred alternative when diagnosing bone injuries and defects in the tissue of the lung and chest. It is also preferred for abdominal pain, showing bone fractures, blood and organ injury. At present, CT is also much less expensive [5].
1.2 Aim

Since the need and usage of CT is widespread, there is a major desire to further improve it. Its major disadvantage is the fact that CT uses X-ray radiation to acquire the images for diagnosis. One of the major factors influencing the quality of the acquired CT images is the amount of radiation. Studies have shown that radiation exposure over certain limits can severely damage the living tissue [3]. Imaging of the beating heart brings additional challenges because of the longer time required to acquire the images of multiple phases of a cardiac cycle. As longer scan time increases the radiation dose, cardiac CT scans are often performed with dose modulation that keeps the radiation exposure low for most phases of the cardiac cycle. This technique, however, results in poorer image quality in the phases corresponding to low doses [6].

1.2. Aim

The aim of this project was to improve the quality of images acquired with cardiac computed tomography by reducing noise levels and increasing image contrast. As the streaking artifacts in many of the project images, caused by metallic implants, also deteriorated the diagnostic quality of the images, their mitigation was included in the overall aim of the project.

1.3 Background

The project was carried out as a master thesis at the technical faculty at Linköping University (LiU). The project was requested by the Division of Cardiovascular Medicine (KVM) at the Institute for Medicine and Health (IMH), an institution within the medical faculty of the university. KVM conducts research within multidisciplinary projects, developing methods and tools for image analysis and visualization. Their main focus areas are within healthcare and medical research. This thesis project was carried out within the imaging branch, focusing on available deep learning methodologies to carry out image enhancement solutions. As KVM has access to large data sets of CT and MRI images, data could easily be obtained at the starting point of the project.

1.4 Research Questions

The following research questions were formulated to focus and further describe the aim of the project:

1. Can an efficient deep learning method be used to enhance the contrast and reduce noise levels in dose modulated CT images?

2. Can a similar method be implemented for mitigating artifacts in dose modulated CT images?

3. Are the enhanced images suitable for diagnosis?

1.5 Delimitations

The length of the project was 20 weeks, and it was therefore important to confine it to a feasible scope. These were the delimitations of the project:

1. All image acquisition was performed before the start of the project, no images were acquired during the project. Therefore, the only focus was to work with post processing of the images.

2. Only cardiac CT images were used and analyzed in the project.
Chapter 2

Theory and Related Work

This chapter covers relevant theory necessary for answering the research questions and understanding the method and results. The chapter covers various aspects of the project, including CT imaging and deep learning. The functionality of CT and its related artifacts are followed by various image quality metrics used to evaluate the results. This is followed by the basics of deep learning, with the theory of how a neural network works. The theory behind convolutional neural networks and generative adversarial networks is then followed. The theory of histogram equalization (HE) is then described, a post-processing technique used for comparison with the proposed method. The chapter concludes with similar projects described in the related works section.

2.1 Computed Tomography

A conventional X-ray scan can obtain 2D-images of any specific area of the body. It does this by emitting an X-ray beam through the material of which a visualization of the interior is requested. On the other side of the material a detector is placed to receive the X-ray beam. Depending on the attenuation of the material that the beam passes through, the beam will have a certain intensity when reaching the detector. In the generated image, higher intensity will correspond to black and lower intensities will correspond to white. For instance, bones have a very high density and will attenuate the X-ray beam greatly, which will therefore be seen as white in the generated image. The lungs that contain mostly air and have very low density will barely attenuate the beam and the resulting image will be black in those areas. The X-ray beam will travel through the entire material of the specified region before the information can be obtained. This means that all layers of the cross section are obtained, all compressed on top of each other into a single 2D-image. Some areas of the image might be impossible to interpret because of this [7].

Computed tomography solves this problem by acquiring information of the interior physiology in 3D, producing many slices (or 2D-images) for the cross-section, giving the information of one specific layer to one specific slice. It acquires information in 3D by emitting X-rays through the body from all angles. The patient is placed into the large CT scanner. A giant ring called a gantry starts spinning around the patient. The gantry contains a tube that will release X-ray beams, and detectors that will measure the amount of radiation that remains after the beams have been attenuated by materials in the body of the patient. The X-ray beams will be able to catch many views of the body from different angles because of the spinning of the gantry. It calculates the density of tiny voxels (3D-pixels) of the entire body. For image processing purposes, the body is divided into a large amount of these voxels. The more voxels the body is divided into, the higher the resolution acquired for the output image. CT calculates the density for each voxel. The detectors then send the received data to a computer that will turn the voxel information into cross-sectional images of the bones and soft
2.1. Computed Tomography

tissues inside the body. CT is more complex than a regular X-ray, but the underlying theory
is the same: lower density areas will be represented by darker areas, and higher density areas
will be lighter. For this reason, CT can outline bones, soft tissue and blood vessels with high
accuracy. An illustration of the structure of a CT scan can be seen in figure 2.1.

Figure 2.1: How the CT acquires images by gantry rotation

2.1.1 The Linear Attenuation Coefficient

What is measured in the CT scan is the linear attenuation coefficient, $\mu$. The attenuation
coefficient represents the degree to which the X-ray intensity is reduced (attenuated) by pass-
ing through a material. The local attenuation is measured in an axial slice of the anatomy
of the patient. The attenuation coefficient of one voxel is translated into the corresponding
grayscale value in the output 2D image slice. The translated output values are normalized to
the attenuation value of water, and are represented on the Hounsfield Units (HU) scale. HU
is also known as the CT value. The CT values are stored as integers in the range of 4096 val-
ues (12 bytes), from -1024 HU to 3071. The scale is based on the attenuation of water which is
given the value of 0 HU. The attenuation of air is -1000, and the values for water and air are
independent of the X-ray spectrum. The formula for converting measured attenuation values
to HU is as follows:

$$HU = 1000 \cdot \frac{\mu - \mu_{water}}{\mu_{water} - \mu_{air}}$$

(2.1)

$\mu$ represents the linear attenuation coefficient of the material in the voxel, $\mu_{water}$ is the linear
attenuation coefficient of water and $\mu_{air}$ is the linear attenuation coefficient of air. When
measuring the values for human tissue, the measurements are greatly affected by the X-ray
spectrum and cannot be given standardized values, only approximations. The lungs are filled
with air and often come close to a CT value of -1000 HU. Fat is also negative, but has a CT
value much closer to 0. Muscles have relatively low positive values, while bones can have a
CT value of up to 2000 HU. Iodine contrast agents can be utilized during scans to improve
visibility of desired areas. Vessels containing the agent can have a CT value in the range of 200
2.1. Computed Tomography

This can be highly useful for generating additional visibility and contrast between the iodine area and its surroundings, as regular tissue often has CT values below 100 (except for bones) \[7\] \[8\].

The linear attenuation coefficient of each voxel is affected by numerous factors, some of them controlled by the clinician managing the CT. The coefficient varies depending on the energy of the photons that crosses the voxel. The photon flux (the number of photons) passing through the X-ray tube will affect the measured photon intensity. If the attenuation is high, the number of photons detected at the X-ray detectors will decrease, and a lower photon intensity is measured. The clinician can control the photon flux and the photon energy, which will affect the output results and the radiation dose to the patient. By increasing the tube current [mA], the photon flux will increase, which will improve image quality and increase radiation dose. The energy of the X-ray can be controlled by the tube voltage [kV]. Increasing the voltage will increase the X-ray energy and make the beam "harder". This will result in lower attenuation rates and increased contrast of the output image. The attenuation coefficient is also affected by the composition and thickness of the voxel \[8\].

2.1.2 Artifacts

In CT, artifacts are distortions or errors in the acquired images from image acquisition techniques. The term is applied to any deviation that occurs between the CT values that are calculated in the reconstructed image and the true attenuation coefficients of the imaged object. Artifacts are more usual in CT than conventional radiographs, as they reconstruct the image from the order of a million different detector measurement points. The reconstruction functionality assumes that the millions of measurements are consistent, leading to measurement errors usually being distinct in the output images. They may appear very different and for different reasons. They often appear as very bright or dark streaks or rings. At worst they could alter the acquired information such a way that a pathological defect or other essential data is lost. Images with artifacts are therefore considered to have significantly lower quality and diagnostic value, sometimes making them unusable for diagnosis \[9\].

In this project, the focus was mainly on one type of artifacts, metal streak artifacts. This was due to the fact that many of the patients that the data was received from suffered from arrhythogenic right ventricular cardiomyopathy (ARVC), an arrhythmia in the right ventricle, for which they had a metallic Implantable Cardioverter Defibrillator (ICD). The ICD showed up as metal streak artifacts in the images. One of these images can be seen in figure 2.2. Here follows the theoretical background of metal streak artifacts.

Metal Streak Artifacts

Metal streak artifacts appear as a result of the high attenuation of metals on or in the test subject. The high attenuation values of the metal are higher than the normal range of CT attenuation values and cannot be correctly handled by the computer. The result is an incomplete attenuation profile. Artifacts arise both on the metal itself and at its edges where it interfaces with lesser attenuating materials. The result is a combination of dark and bright streaks.

Metal objects can often lead to additional artifacts due to its very high density, such as beam hardening artifacts and partial averaging volume artifacts. The initial approach to avoid artifacts caused by metallic objects is to remove the metal from the patient. If the metal is impossible to remove, it can sometimes be possible to manipulate the gantry rotation to angle it to exclude the metal areas. However, this can in many situations be too difficult, as in the case when the patient has an ICD and the metal object is positioned exactly in the region of interest. Another possible solution is to increase the voltage. This has the potential to allow
the X-ray beam to penetrate certain objects, and by additionally using thinner CT sections (ac-
quiring smaller voxel volumes), it can reduce the partial volume artifacts considerably. There
also exist various interpolation techniques that can be used in the reconstruction software to
reduce the streaking caused by overranging. These techniques are often adept at removing
streaks that are distant from the metal, but poor with those that are closer to the source, which
is often most essential for diagnosis. It can also be fruitful to utilize beam hardening correc-
tion software to minimize additional beam hardening artifacts caused by the metal [9] [10].

Figure 2.2: A cardiac CT image with an ICD from the data set. The artifact is apparent and reduces
image quality considerably.

2.2 Image Quality Metrics

Visual evaluation of the output images is important for initial assessment of the image quality,
but it is not enough. Visual inspection is difficult to quantify and is highly subjective. Some
metrics must be used to quantify the results.

2.2.1 Signal to Noise Ratio

Signal to noise ratio (SNR) is a popular metric used for evaluating the quality of an image
based on its inherent noise levels. In signal theory it is simply defined as the ratio between
the strength of the signal and the approximated noise.

\[ \text{SNR} = \frac{S}{N} \]  

In imaging, it can be calculated in different ways. In this project, the signal is acquired by
first making a region of interest (ROI) in the image. The aim of this is to measure the signal
strength of the information within this area. This is done by taking the average value of all
its pixel values.

\[ S = \mu(\text{ROI}) \]  

In a gray scaled CT image, each pixel is represented by a normalized value (often between
0 and 1, or between 0 and 255). The lowest value corresponds to black, the highest value
corresponds to white, and the values between them to different shades of gray. The noise is
approximated from the image’s background, which contains information that is uninteresting
for the observations. The noise is calculated as the standard deviation of the background.

\[ N = \sigma(\text{background}) \]  

By inserting equation 2.3 and equation 2.4 into equation 2.2, SNR is defined with the follow-
ing formula [11]:

\[ \text{SNR} = \frac{\mu(\text{ROI})}{\sigma(\text{background})} \]
2.2.2 Contrast to Noise Ratio

Even if the SNR of the image is high, the image will have low visibility without contrast. If there is low contrast, a majority of the pixels will have closely set intensity values, making it more difficult to perceive differences within the image. Contrast to noise ratio (CNR) is a metric utilized to measure the contrast within the image compared to the noise levels.

\[
\text{CNR} = \frac{C}{N}
\]  

(2.6)

To establish the CNR, the contrast and the noise must first be defined and calculated. Contrast is the difference in intensity between two signals. In an image, the signals are seen as the pixel intensities. A ROI must be established, containing the requested area to measure. If the aim is to measure the contrast between two different areas of the image, two different ROIs must be defined for these areas. Each of the signals is estimated as the mean pixel value in their respective ROI. The contrast is the difference between those two values.

\[
C = S_1 - S_2
\]

\[
C = \mu(ROI_1) - \mu(ROI_2)
\]  

(2.7)

However, the interest is often in only one specific area of the image, and in that case only one ROI must be defined. The background of the image, which is everything not included in the ROI, is passed as the second signal.

\[
C = \mu(ROI) - \mu(background)
\]  

(2.8)

To acquire a mathematical value for the noise in both of the aforementioned cases, a background region is defined in the remainder of the image where there are no ROIs. This is performed in the same manner as for SNR, equation 2.4. In the background area of the image there is no interest to calculate signal intensity, since there is no signal of interest. The noise is then defined as the standard deviation in this area. The CNR is the ratio between the estimated contrast and noise as seen in equation 2.6. In this project, the second method for calculating contrast has been used, equation 2.8. The values from equation 2.4 and 2.8 are therefore inserted into equation 2.6:

\[
\text{CNR} = \frac{\mu(ROI) - \mu(background)}{\sigma(background)}
\]  

(2.9)

2.3 Deep Learning

Artificial intelligence (AI) is the creation of software systems with a behaviour that tries to imitate human thinking. The actions of the system are often managed by a ready-made set of rules. Machine learning is a subset of artificial intelligence. The main difference is that a machine learning system focuses on techniques that enable learning from large amounts of data. It is not explicitly programmed or instructed how to solve a problem, but must learn this by itself. Artificial intelligence outside the scope of machine learning is ready to be used directly after implementation, and does not require training.

As machine learning is a subset of AI, deep learning is a subset of machine learning. It was initially inspired by the structure and function of the brain, specifically the connections between the many neurons. However, the deep learning techniques developed today are only loosely based on this. The deep learning neural networks are composed of artificial neurons divided into many layers. Each layer specializes in learning to recognize a specific feature of the data. The layering results in depth, which defines deep learning and gives it its name.
2.3. Neural Networks

The elementary unit of the conventional neural network is called a neuron (figure 2.3). It can be seen as a small system or function that receives input, performs a few calculations and then outputs a value.

The input is data from neurons in a previous layer, or from the original input data if the neuron is in the first layer. The input $X_i$ is multiplied by a value called a weight $W_i$ that is assigned to the vector between the source and the current neuron. All the products from the previous layer are summed. A value called bias $b$ is then added. The result is the input to a so called non-linear activation function, noted by the $f$ in figure 2.3. The output from the activation function, also called the activation, is what the neuron outputs and sends to the neurons in the next layer. The mathematical expression is then (as seen in figure 2.3):

$$f\left(\sum_{i=1}^{n} (W_iX_i) + b\right)$$

These neurons then build up the neural network as can be seen in figure 2.4. It consists of an input layer (blue in figure 2.3) that takes in the original data as input, an output layer (yellow in figure 2.3) that yields the resulting output and an arbitrary number of hidden layers in between. All neurons in one layer are connected to all the neurons in the neighbouring layers. The weights and biases are the parameters of the network that change throughout training.

By finding proper values for these parameters, the model can be trained to recognize certain
patterns. After the network has gone through its training, the values have hopefully reached optimum values for performing the task that the network is designed for.

The weights are used for activation between the layers. If some input values are of interest in a certain layer (maybe they make out a pattern in an image) then those values could be multiplied with high valued weights while the other input values are multiplied by low weights to enhance the visibility of the pattern of interest. Changing the weights will change the shape of the non-linear activation function. The bias is added to the neuron to make it possible to shift the entire activation function along the x-axis which can be very useful. E.g. by setting the bias to a large negative number, smaller output (smaller than the bias) from the first part of equation 2.10, \[ \sum_{i=1}^{n} (W_iX_i) \], will be zero-valued or negative. If a sufficiently high number is output from \[ \sum_{i=1}^{n} (W_iX_i) \], a positive value will be the input to the activation function and the network will interpret this value to be meaningfully active [13].

The results from the input values, weights and the bias are passed through the non-linear activation function in every neuron (2.10). The purpose of it is to make sure that the neurons are not linear systems that manage input and output values in a purely linear translation. The reason is that the task of the neural network rarely is of a linear nature, otherwise simpler methodologies with much less complexity could be implemented instead. The aim of the network is most often to find correlations and mappings in highly complex circumstances, with multiple features and patterns that may or may not yield information of the correct mapping to the output. The network must not only find the numerous features but also figure out which of them are of any interest and which are not. By only using linear connections within the network it would be impossible to find all the different mappings. One possible perception of the hidden layers is that they distort the input in a non-linear fashion, making the groupings linearly distinguishable in the last layer.

2.3.2 Supervised vs Unsupervised Learning

When training a neural network model, supervised learning is utilized in most cases. This is possible when ground truth exists: training data paired with corresponding labels. The label contains the information of the correct class or segmentation of the input data. When the training begins, the weights and biases are often initialized to random values or set to zero. When the data runs through the network, it can see directly how well it performed during one iteration by looking at the corresponding label of the data [13] [14].

It is possible to perform training of the network even when no ground truth is available, and it is then called unsupervised learning. If ground truth is available it is the typically the preferred alternative as it is more straightforward. However, it can be quite expensive to classify data and produce labels, which is a task that must often be performed manually. Unsupervised learning is more complex but can be a cheaper alternative if there is no labelled data available. The network is in this situation handed a data set without instructions. It is not instructed what to do with the data set or what the desired output is. The network simply attempts to find patterns in the data by locating features that might be useful and analyzing their structure. There exist a few popular methods for this.

- Clustering is one method, were the network realizes that many data samples belong together in clusters based on a number of features. The network can then hopefully group most of the data into different clusters.
- With anomaly detection, the network looks for patterns of outliers within certain features. It finds an “anomaly” if a majority of the data is similar in a certain aspect, but a few samples are not.
• By identifying that certain features of a data sample correlate with other features, the network can comprehend association. The network can then look at a certain feature of a data point and from that information predict specific other features of that data point.

Semi-supervised learning is another concept that lies somewhere between supervised and unsupervised learning. It is defined as learning from a small set of labeled data. E.g. if an expert is needed to perform the manual labelling, it can be very costly. An option is then to let the expert label only a small set of the data to not spend too much time. This small set can then be used for training. A generative adversarial network (described further in section 2.5) is sometimes described to train with semi-supervised learning. It consists of two networks, one that generates new data and one that classifies between the generated data and a true data set. The classifying network makes use of the labelled true data set and the generated data (which is labelled as ‘generated’) to learn to distinguish between the two. It is possible to use a labelled data set of smaller size, in which case the network is said to be performing semi-supervised learning (even though more data is most often recommended). In this manner, the network can benefit from the smaller proportion of labelled data, which can be a great advantage over a fully unsupervised model.

2.3.3 The Cost Function

A cost function is utilized to calculate a value representing the performance of the network. The cost function is used during training of a neural network to express how well the network’s prediction is at this point in the training. It is a mathematical function that outputs the cost of the prediction, which measures the difference between the prediction and the true value. If the result from the cost function is high (large difference between output and correct value), the network knows that it must make large changes to improve. The network improves by changing the values of the weights and biases. It knows in which direction to change the parameters based on the slope within the multivariate cost function. By finding the direction of the steepest slope, where the function moves upwards the fastest, the direction for minimizing the cost is the negative direction of that. Finding the global minimum of the multivariate cost function is the ultimate aim of training the network, although finding a low local minimum can in many cases yield quite good enough results. However, a regular problem that can occur is that the network finds a local minima and stops improving its parameters. The step size or learning rate is a hyperparameter of the network that defines a value that the network must adjust during training. It is the magnitude that instructs the network of how much it will travel along the downwards slope of the cost function, and thereby how much the network parameters should be changed for one iteration of the training. Setting the correct learning rate is vital. If it is too low the network has a higher risk of getting stuck in a local minimum. If it is too large the network instead has a higher risk of overshooting and missing the minima. Setting an adequate fixed learning rate is often difficult, which is why in many cases the learning rate is changed throughout the training. Many variations of cost functions exist.

An often used cost function is the mean absolute error (MAE) cost function, which simply takes the absolute value of the difference between the prediction and the correct value.

\[
MAE = \frac{1}{n} \sum_{i=1}^{n} | y_i - \hat{y}_i |
\]  

(2.11)

Where \( y \) is the predicted value and \( \hat{y} \) is the true value. MAE outputs the average of the actual error between the prediction and the true value and nothing else. The MAE is a linear function, meaning that all the individual differences are weighted equally in the average.
Another well used cost function is the mean squared error (MSE) cost function which takes the difference between the prediction and the correct value and squares it.

\[
MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2
\]  

(2.12)

The MSE will output higher costs than MAE if the difference is greater than 1, and it will increase more quickly. When the difference increases, the MSE punishes the network to a greater degree by informing it that it is performing very poorly. The network must then make greater changes compared to if the MAE was used. This can be useful if the network outputs unexpected values that it should not, so that the network can be punished to a high degree and learn to not repeat that behaviour. If the difference is below 1, the MSE cost will be quite low, lower than the MAE cost, and the network is instructed to only make small changes to its parameters. A disadvantage with the MSE is that the network can receive a quite high cost even though it is not performing very poorly, and it is susceptible to outliers [12] [15].

2.3.4 Gradient Descent and Backpropagation

The steepest slope of the cost function, the direction of the steepest increase, is also known as the function gradient. By moving in the opposite direction of the gradient, taking the negative gradient, the direction to move to acquire the steepest decrease is acquired. The size of the gradient indicates how steep the steepest slope is. The negative gradient is essentially a vector with values, one for each parameter of the network. The magnitude of the value represent how much the different weights and biases should be shifted in the next step of the training, and the sign of the value tells the network if the parameter should be increased or decreased. A larger magnitude corresponding to a parameter means that it is more important that that parameter is adjusted. Changing that specific weight or bias will decrease the cost more than changing another parameter with a smaller magnitude in the gradient vector. Hence, the cost is more sensitive to alterations of the weights and biases with larger magnitudes in the gradient [13].

The gradient is calculated in the backpropagation process, the core algorithm for how the network learns. The process begins in the neurons in the last layer of the network, the output layer. The cost function is used to calculate how large the error is and the magnitude of the alterations that should be made. The results from the cost function affects the output layer neurons, they must adjust to minimize the cost. The output from the neurons are not only affected by their respective weights and biases, but also by the output from the neurons in the previous layer. These output values must also be adjusted to optimize the output from the output layer neurons and minimize the cost. A chain reaction is started that propagates through the entire network, from the output layer neurons to the input layer neurons. Backpropagation through the entire network is performed once for every input data point in the training set, and a temporary gradient vector is calculated each time. Once the backpropagation has been performed on the entire training set, all the temporary gradient vectors are used to calculate an average gradient vector, the actual gradient. The gradient is then used to perform one step of gradient descent. A typical training is usually tens of thousands of steps, or more, and it becomes apparent why deep neural networks require substantial computational power [13].

**Stochastic Gradient Descent**

For a true gradient descent step, the network should use every data point from the training set to train on for every iteration of the training. Since this in most cases will require a lot of computational power (depending of course on the size and structure of the input), stochastic
2.4 Convolutional Neural Networks

The conventional neural network looks at an image by taking all of its pixels as one input vector. In this vector, a pixel is only connected to at most two other pixels, the one in the previous position of the vector and the one after. This means that the network’s spatial information of the image, how several pixels together can form different features or patterns, is poor. The convolutional neural network (CNN) has proven fit for increased spatial understanding and improved image processing. Its structure is built upon the original neural network, with added functionality applied to the image before the very last layer, called the fully connected layer, which is often similar to the conventional neural network. The CNN takes in an image as a matrix instead of a vector, which enables the network to acquire an increased amount of spatial information. The network then performs convolution on the image, which is the process of acquiring feature maps. A filter or kernel, which is a small matrix, runs over the image and pauses at certain intervals. The size of the interval is called the stride. At every stride the dot product between the filter and the current position of the image is calculated. The final result of the convolution is a feature map with a smaller width and height than the original image. The feature map contains features and patterns that let the network learn what is of interest in the images. One convolution operation between one kernel and one image will output one feature map, so running multiple kernels on one input image will result in multiple feature maps. To avoid too heavy computations and decrease memory usage, the feature maps are typically pooled, or subsampled. The most common technique is max pooling, where a small window runs over the feature map and only outputs the maximum value in the window at the current position. The maximum number will have most significance, and in this manner required computational power can be reduced while maintaining the most essential information. The processes of convolution and pooling in a CNN can together be called a convolutional layer, and a CNN usually consists of an arbitrary number of these. After each layer, the number of feature maps will increase while their size is reduced. Low level features such as surrounding edges are typically recognized in the first layers, followed by more and more detailed features in the following layers, corresponding to the higher number of feature maps of smaller size. The final feature maps will be flattened into vectors so that they can be used as input to a fully connected layer, where they are processed in the same manner as in a conventional neural network. The structure of a classification CNN is depicted in figure 2.5 [13] [16].

2.5 Generative Adversarial Networks

A general disadvantage of the earlier (before 2014) image related neural networks was the inability to generate images that were passable as non-generated, their typical uses were to classify them or to perform image segmentation. In 2014, Ian Goodfellow and his team imple-
2.5. Generative Adversarial Networks

A GAN has the ability to generate random images of requested motives by receiving random noise. It consists of two separate neural networks, often convolutional neural networks in deep convolutional GANs, that work as each others’ adversaries. The first network, the generator, receives some random noise and uses it to create an output image. The discriminator in a GAN is implemented as a binary classifier. It receives data from a training set consisting of unaltered data, and the output from the generator. Its task is to learn how to correctly classify them. There will always only be two classes: unaltered data and the data generated by the generator. The discriminator learns as a conventional classifying neural network, with a cost function. If it classifies an image incorrectly, the cost will be high and vice versa. When the cost is high the discriminator is “punished”, meaning that it is instructed to perform major changes in its parameters because it is performing poorly. As it works to minimize the cost from the cost function, the generator attempts to increase it. If the discriminator cost is high, that means that the generator is performing well, it succeeds in fooling the discriminator to classify images incorrectly. If the discriminator cost is low, it means that the discriminator is performing well and succeeds in classifying the images. That also means that the generator is performing poorly, the generated images are too unlike the images from the real dataset. It is instructed to perform major changes to its parameters to achieve better results. The structure of the GAN is built on this, a minimax two-player game between the generator and the discriminator, both trying to affect the cost of the discriminator. The two networks work against each other in this manner, as each others adversaries, forcing each other to improve.

The discriminator utilizes backpropagation to learn and update its parameters. In this manner, it learns how to recognize features: which features occur in an image from the real dataset and which occur in a generated image. Backpropagation is also used to update the parameters of the generator. The basic structure of the conventional GAN can be seen in figure 2.6.

2.5.1 Cycle-Consistent Generative Adversarial Networks

In a conventional GAN, the generator usually receives random data from which it produces a requested output. By feeding the generator random data points, it can produce different looking output of the same motive. In this manner, images of random animals, buildings, faces, maps, etc. can be generated. The purpose of the Cycle-consistent generative adversarial network (or CycleGAN) is not to output random images of a certain motive, but rather to
2.5. Generative Adversarial Networks

Figure 2.6: The structure of a conventional GAN.

attain the same image from a different motive, e.g., taking an image of a horse as input and generating the same image where the horse has been transformed into a zebra. The terminology that will be used from here on is that the CycleGAN translates images from domain A to domain B, also known as image-to-image translation. The network does not only translate images from domain A to domain B, but also in the opposite direction [19].

The problem with generating images from one domain to another is the need for paired image data for the training set, to make it possible to perform supervised training. An example would be that there are two data sets, and for every image of a horse in the first data set there is an identical image in the second, with the only exception of the stripes on the animal. An example of an image pair such as this is depicted in figure 2.7. The complication is the fact that image pairs such as these rarely exist naturally. If they did, the paired data could be used as ground truth when training a network, and used to tell the network how close the output is to the correct output. Without it, no supervised learning can be achieved.

Figure 2.7: A pair of images from two different domains. An example of paired data [19].
The CycleGAN is a technique that brings a solution to this problem. An illustration of the network can be seen in figure 2.8. It consists of two different GANs, each consisting of a generator network and a discriminator network. The first generator, $G_{AB}$, receives an unaltered domain A image, $\text{original}_A$, and translates it to a domain B image, $\text{translated}_B$. $\text{translated}_B$ is then given to the domain B discriminator, $D_B$. $D_B$ also receives images from the domain B training set to work with, $\text{original}_B$, and its task is as in a conventional GAN to classify unaltered and generated images ($\text{original}_B$ and $\text{translated}_B$). The cost function from $D_B$ is used to improve both $D_B$ and $G_{AB}$. This cost says how well the network can translate an image from domain A to domain B, the probability that the generated $\text{translated}_B$ belongs in domain B.

However, the cost function of $D_B$ says nothing of how similar $\text{original}_A$ is to $\text{translated}_B$, and it is of high importance that $\text{translated}_B$ is a domain B replica of $\text{original}_A$. This is solved by giving $\text{translated}_B$ to the second generator network, $G_{BA}$, that translates it back to a domain A image, $\hat{\text{translated}}_A$. $\hat{\text{translated}}_A$ is then compared to $\text{original}_A$, and if both translations were successful the two images should be very similar. The difference between them is used to calculate a cycle-consistency cost, which says if the network is translating images with their structural integrity intact. If the images are similar enough, then $\text{translated}_B$ should truly be a domain B replica of $\text{original}_A$. A more detailed illustration of the network and how the cycle-consistency cost works is depicted in figure 2.9. In this manner the CycleGAN learns to translate images back and forth between two domains [19].

![Figure 2.8: A visualization of the CycleGAN.](image1)

![Figure 2.9: A visualization of how the cycle-consistency cost works in the CycleGAN.](image2)
2.6 Histogram Equalization

A histogram is a diagram illustrating the frequencies of occurrence of the different values being measured. The histogram of a gray scaled 8-bit image will thus represent the occurrences of each of the values between 0 and 255, as those values would represent the pixel intensities, the gray levels. 0 represents black, 255 represents white and the values in between represents different shades of gray. In figure 2.10, a gray scaled image is depicted together with its corresponding histogram. Much information about the image can be obtained from the histogram.

Figure 2.10: A gray scaled image and its corresponding histogram.

If the histogram is narrow, it can be concluded that the image is poorly visible because the amount of gray levels in the image that can be used to visualize the contents are generally low. This also means that the contrast must be generally low, since the difference between high and low pixel intensities is small. A histogram with a wide distribution means that a majority of the gray levels are included in the image, which should correspond to a generally high image contrast and content visibility [20].

Histogram equalization (HE) is based on the principle that if the majority of the pixels in a gray scaled image are within a small range of the available intensities, the image will have low contrast. By expanding the dynamic range of the pixel values, the overall contrast of the image can be increased, and thus also the image quality. The lighter pixels will become even lighter and the darker pixels will be further darkened. The gray scale of the image is adjusted by mapping the input image histogram onto a uniform histogram. Images with nonuniform histograms often have small dynamic ranges, which can be increased by transforming the histogram from nonuniform to uniform. Figure 2.11 shows how the overall contrast and visibility of the image is improved with HE [20].

In figure 2.12 the new image and the new histogram can be seen. The new histogram is uniform and has more evenly distributed values than the original in figure 2.10.
2.6. Histogram Equalization

Figure 2.11: A gray scaled image when it is processed with HE. The overall contrast is increased.

Figure 2.12: The image after HE and its corresponding histogram.

2.6.1 Adaptive Histogram Equalization

If there are regions within the image that are considerably lighter or darker than the remaining pixel values, the conventional HE will not enhance those regions sufficiently. A solution to this is the adaptive histogram equalization (AHE). While the conventional HE obtains and alters the histogram for the entire image, AHE looks at several distinct areas of the image and obtains their histogram. It then redistributes the values within the local region to improve contrast and enhance the visibility of the edges there. It does this in the same manner as conventional HE [21].

When the contrast within a neighbourhood is low, the pixel intensities within that region will be quite homogeneous. The histogram will then have a narrow peak with the majority
of the pixel values. This narrow range of pixel values will be mapped to the whole range of the output. Hence, AHE generally overamplifies contrast in the homogeneous (or near-constant) areas of the image. As a result, small amounts of noise may also be overamplified in these regions. A contrast limitation technique has been added to the AHE to prevent this. The contrast amplification is then limited, to mitigate the noise amplification. In contrast limited adaptive histogram equalization (CLAHE) the contrast amplification is limited to a predefined value (the so-called clip limit). The values in the histogram that exceeds this limit are often redistributed equally among histogram values. In figure 2.13 the image has been processed with CLAHE and is depicted alongside its corresponding histogram [22]. In this project, HE and CLAHE were used on a test set of images to compare its results with the results from the implemented method. Since all these methods aim to increase contrast and improve image quality, the comparisons were of interest.

2.7 Related Work

In October 2017, a team from Korea published an article where they revealed that they had experimented with removing artifacts from low-dose X-ray CT scans by using a CNN in combination with directional wavelets [23]. The results were promising, and they far surpassed earlier techniques using model-based iterative reconstruction (MBIR). Their conclusion was therefore that their algorithm would open possibilities in an area of low-dose CT research. The article was of high relevance when performing this project, as it had a similar approach to the problem with artifacts in CT images [23].

In another study from 2017 by Wolterink et al. [24], a conventional GAN was used to increase image quality from CT scans. More specifically, the purpose of the project was to reduce noise levels generated from low-dose CT image acquisition. The approach was to develop a GAN consisting of a CNN generator and a CNN discriminator. With these two adversaries an estimation of the difference between routine-dose CT images and low-dose CT images was made to make it possible to reduce noise. The generator network was trained to transform low-dose CT images into routine-dose CT images using voxelwise cost minimization. The generated
images were then sent to the adversarial discriminator network which tried to classify these generated images and true routine-dose CT images. The performance of the discriminator was measured with a cost function, which was used as an adversarial cost for the generator. Three different cost calculations were used for training the networks.

- Only voxelwise cost
- Voxelwise cost combined with adversarial cost
- Only adversarial cost

Here meaning that the adversarial cost was the result from the cost function of the discriminator network. The results from comparing the three strategies showed that the first technique, only using voxelwise cost, gave the highest peak signal-to-noise ratio [24].
Chapter 3

Method

In this chapter, the method used for finding the answers for the research questions is described. The project consisted of: (a) classification of available images into training, validation and test sets for the neural networks, (b) implementation of these networks, and (c) evaluation of obtained results.

3.1 Image Acquisition

The CT images used in this thesis project were acquired in 2016, before the project initiation. Approximately 130 000 2D cardiac CT images were used, sliced from the original 3D-images in cardiac short-axis direction from 55 patients. Twenty 3D-images were available for each patient representing 20 consecutive phases of the cardiac cycle, resulting in a total of $55 \times 20 = 1100$ 3D-images. Of the 55 patients, 19 had ARVC. For this, they had an ICD which resulted in metal streak artifacts in the corresponding CT images. The patients with arrhythmia will be henceforth referred to as the A-patients, and the images acquired from them, with metal streak artifacts, will be called A-images. The remaining 36 patients either had suspected Coronary Artery Disease (CAD) or a Transcatheter Aortic Valve Implant (TAVI). The images from these patients did not have metal streak artifacts. These patients will be referred to as the R-patients, and the images acquired from them will be called R-images.

All images were acquired over 3-4 cardiac cycles (depending on the heart rate). The R-images were acquired with a dual source CT scanner (Siemens Force, Siemens, Germany). Retrospective image acquisition was utilized with ECG-triggered dose modulation. Hence, the radiation dose was lowered during the cardiac phases that were predicted to be suboptimal for image reconstruction [6]. The 2D image resolution of the image slices was 512×512 pixels, and the reconstructed 3D resolution for each slice was 0.35×0.35×0.25 mm³. Further reformation to short-axis views and isotropic resampling resulted in a final resolution of 1×1×1 mm³ voxels. Ethical approval to use the R-images was given by the regional Ethical Review Board (ERB) at Linköping University Hospital with the condition that no patient information would be used, only the images.

A dual source CT scanner (Siemens Force, Siemens, Forchim, Germany) was used to obtain the A-images. Retrospective ECG-triggered dose modulation was used for these images as well, in a spiral manner. Reformatting was performed to obtain short-axis images with a final resolution of 1×1×1 mm³ voxels. Written informed consent for usage of the images was received by all A-patients. A more detailed description of the image acquisition is given in the article by Gupta et al. [25], the study for which the images were originally obtained.
3.2 Classification of Images

To perform training and testing with the networks, it was necessary to divide the data into subsets. Furthermore, these sets had to be divided into two separate classes, one with low quality images and one with high. These classes would be fed to the networks as the two image domains, so that the network could learn how to translate back and forth between them.

12 A-patients formed the A-training set and the other 7 became the A-test set. The R-training set consisted of 25 of the R-patients, and the remaining 11 formed the R-test set. The R-images were utilized for training the network for image quality enhancement, increasing contrast and reducing noise, while the A-images were utilized for training the network to minimize the presence of image artifacts.

The R-training set was obtained by classifying all the R-images into high and low quality bins. Initially, the quality was to be determined from two metrics, SNR and CNR using a ROI that was defined as a rectangle including both cardiac ventricles. However, this gave no clear clusters of images and a decision was made to only focus on one parameter. High contrast within the ventricles was finally deemed of higher importance than noise levels, and the only measured parameter was the CNR. One arbitrary 2D-slice in the mid-ventricular region of each 3D-image was retrieved. The purpose was to acquire one slice from each 3D-image that distinctly contained both ventricles. The middle slice was then tested and evaluated for image classification.

Image quality is a highly subjective matter, but it is generally agreed upon that it cannot be measured from a single metric and is dependant on a diverse set of variables. Therefore, the classification was not based on the CNR alone. It was rather used as a first step for classification. The resulting classes were then analyzed manually to make sure that no apparent cases of incorrect classification had occurred.

Classifying every single slice of each 3D-image would entail a tedious and time consuming process, especially with manual control of the classification afterwards. This is why only the middle slice was used for each 3D-image. It was also deemed unnecessary to classify every 2D-slice as there was an assumption that the contrast was constant within one 3D-image and time frame. Hence, if the middle slice of a 3D-image was classified as high quality, the entire 3D-image and all of its slices were automatically classified as high quality as well, and vice versa.

The CNR was calculated for the middle slice images. Two different classifications were then performed. The first method classified the top 30% of the images, the 30% with highest CNR values, as high quality. The bottom 30%, the 30% of the images with lowest CNR, were classified as low quality. The middle portion of the images, those that were not deemed to belong to either the top portion or the bottom portion, were not used as training data for the network. In the second classification method the network was trained with the middle portion of the data included. In this classification the top 30% of the images were classified as high quality and the remaining 70% were classified as low quality. Both methods were utilized and their results were compared.

3.3 CycleGAN

As explained in section 2.5.1, the CycleGAN consists of two separate GANs with two cost functions each. Instead of using horses and zebras as in the example described, the A-domain of the network used in this project consisted of low quality images and the B-domain consisted of high quality images. The purpose was to train the network to learn how to translate images between the two domains, that is, from low quality images to high quality images.
3.4 Evaluation of Results

Two separate CycleGAN models were created, one with the aim of primarily increasing contrast and secondarily to reduce noise levels, and one with the aim to reduce the appearance of artifacts. They were named the enhancement network and the artifact reduction network and will from here on be referred to as such. The main difference of the implementation of the two networks was the input training images. The enhancement network was given low quality training R-images as domain A images, and high quality training R-images as domain B images. The first implementation with classification of the training data that excluded the middle portion will from here on be referred to as enhancement network 1. The second implementation with classification of the training data that included the middle portion of the training data will from here on be referred to as enhancement network 2. The artifact reduction network was given training A-images as domain A images and high quality training R-images as domain B images. All domain data sets were then divided into a separate training set and a validation set with the ratio 70/30.

The cost functions used in the model was the MAE cost function for the cycle-consistency cost, and the MSE cost function for the remaining costs used within the GAN (unmodified from the original implementation of the CycleGAN). An attempt to increase the CNR and SNR of the output images from the enhancement network (enhancement network 2) was made with a modification of one of the cost functions. As in the GAN implemented in section 2.7, the idea was to make an attempt to give the network further instructions. The network was not only to translate images from domain A to domain B, but also to directly improve the image quality. The cost function of the domain B discriminator was adjusted. It was previously learning with a conventional MSE cost function, but the CNR and SNR difference between the translated image and its corresponding original was added to increase the cost when outputting low CNR and SNR differences. The hope was that the network would be forced to increase the CNR and SNR to reduce the cost. This implementation will from here on be referred to as enhancement network 3.

3.4 Evaluation of Results

3.4.1 Image Quality Metrics

During the training of all networks, the model containing all parameters (weights and biases) was saved after each epoch. The networks were trained during 80 epochs, so 80 different models were saved for each training. After training of the enhancement network, the validation set of R-images was used to generate processed images from each epoch. For each image, the SNR and CNR were calculated as described in section 2.2. The difference between the SNR and CNR of the processed image and the original image was then calculated.

\[
\text{CNR}_{\text{difference}} = \text{CNR}_{\text{processed}} - \text{CNR}_{\text{original}}
\]

\[
\text{SNR}_{\text{difference}} = \text{SNR}_{\text{processed}} - \text{SNR}_{\text{original}}
\]

The percentage difference was then calculated to quantify the difference by comparing it to the original value,

\[
\text{CNR}_{\text{percentagedifference}} = \frac{\text{CNR}_{\text{difference}}}{\text{CNR}_{\text{original}}}
\]

\[
\text{SNR}_{\text{percentagedifference}} = \frac{\text{SNR}_{\text{difference}}}{\text{SNR}_{\text{original}}}
\]
The mean value from each epoch was then calculated. Towards the end of the project, when all the network’s hyperparameters had been established, a randomized test set of 3000 R-images was used to perform the final evaluation of the enhancement network. The test set consisted of images of both high and low quality that were unused during the entire development phase, to make sure that the network was completely unfamiliar with them. SNR and CNR was calculated in the same manner as for the validation set, but the mean values were only calculated for every fifth epoch. These values were plotted against the epochs throughout the training and the plot trends were acquired. HE and CLAHE were also performed on the test images, and the SNR and CNR average differences were calculated for these as well. The results from HE and CLAHE were compared with the results of the different implementations of the enhancement network.

The costs of the network were also recorded throughout the training, but it was deemed of less importance than the quality metrics, as the main purpose of the training was to improve image quality. Minimizing the network cost is the main purpose when performing classification or regression rather than data generation. It is of course useful to see how the network costs are decreasing throughout the training, but it has no direct correlation with the results. It was mainly used to confirm that the network was assessing its own performance as improving, and that no overfitting was happening, as the validation cost and the training cost did not diverge.

### 3.4.2 Qualitative Test

CNR and SNR give two measurements trying to define the image quality, but it is of great relevance to complement them with a visual inspection. A qualitative test was created to acquire more results of the image quality of the processed images. The test was done by three medical students and five clinicians experienced with CT images. The images used in the test were output images from enhancement network 2 and the artifact reduction network. The aims of the test were multiple, and it was therefore divided into three parts.

In the first part, the evaluators were asked to look at 30 single images and answer if they were original CT images or processed with the implemented software of the project. The aim was to see if there were considerable overall visual differences between the original images and the processed images. If it is always possible to say if an image is processed, it could be a result of the network producing images with visual flaws.

In the second part, the evaluators were presented with 42 image pairs. 28 of the image pairs were R-images and the remaining 14 were A-images. The pairs consisted of an original image and its corresponding processed version generated by enhancement network 2 and the artifact network, and it was known to the evaluator which was which. The evaluators were asked to grade the processed images compared to the original based on three different criteria: contrast, noise and artifacts. The grading scale was 1-5 for each criteria:

1. Processed image considerably worse than the original.
2. Processed image slightly worse than the original.
3. No notable differences between the images.
4. Processed image slightly better than the original.
5. Processed image considerably better than the original.

The test was a method for evaluating the effects the implemented method had on artifacts. The aim was also to acquire additional results of the difference in contrast and noise. Coefficients of variation were calculated for the grades of the different images to assess the degree
of agreement within the test group.

In the third part of the test, the evaluators were given 40 image pairs. The image pairs once again consisted of an original image and its corresponding processed version, processed by enhancement network 2, but in this test the evaluators were not informed of which image was which. They were then asked to say which of the two images had higher diagnostic quality. The aim of the test was to analyze if the processed images were preferred for diagnosis or not, without any bias that might arise from knowing the identities of the images. Only R-images were used in this test as the artifact reduction network was deemed to output images of too poor quality to make this test necessary.
Chapter 4

Results

This chapter presents the results from the project including the image quality metrics calculated for the different implementations of the enhancement network. No metrics were calculated for the artifact reduction network, but original images and their processed equivalents are presented for visual illustrations of the results. Images from the enhancement network are also included, as well as images processed with HE and CLAHE for comparison. The chapter concludes with the results from the qualitative test.

4.1 Image Quality Metrics

Enhancement Network 1

The initial approach was to only use the top and bottom 30% of the training dataset. The 30% of the images with the highest CNR were classified to domain B and the 30% of the images with the lowest CNR were classified to domain A. The middle portion, the remaining 40% of the training dataset, was not utilized. The calculated results are visualized in figure 4.1.

Figure 4.1: The results from enhancement network 1. Epochs are presented on the x-axis and the SNR and CNR values are presented on the y-axis. Every 5th epoch is included from 1 to 80 (including 80).
Figures 4.2 and 4.3 include a selection of results from enhancement network 1.

Figure 4.2: Test images from enhancement network 1. Original images are displayed to the left and their corresponding processed versions are on the right.
Figure 4.3: Test images from enhancement network 1. Original images are displayed to the left and their corresponding processed versions are on the right.
Enhancement Network 2

The second approach was to utilize the entire training dataset. The 30% of the images with the highest CNR were classified to domain B and the remaining 70% of the images were classified to domain A. The calculated results are shown in figure 4.4.

Figure 4.4: The results from enhancement network 2. Epochs are presented on the x-axis and the SNR and CNR values are presented on the y-axis. Every 5th epoch is included from 1 to 80 (including 80).

Figures 4.5 and 4.6 include a selection of results from enhancement network 2.
Figure 4.5: Test images from enhancement network 2. Original images are displayed to the left and their corresponding processed versions are on the right.
Figure 4.6: Test images from enhancement network 2. Original images are displayed to the left and their corresponding processed versions are on the right.
Enhancement Network 3

Adding functionality to enhancement network 2, using the entire dataset, a third approach was to add CNR and SNR percentage difference to the domain B discriminator cost function. The aim was to force the network to increase the CNR and SNR to lower its cost. The calculated results are shown in figure 4.7.

![Graph of Average Epoch Values](image)

**SNR Evolution**

- **Difference (%)**
  - 0
  - 5
  - 10
  - 15
  - 20

- **Epochs**
  - 0
  - 10
  - 20
  - 30
  - 40
  - 50
  - 60
  - 70
  - 80

**CNR Evolution**

- **Difference (%)**
  - -50
  - -40
  - -30
  - -20
  - -10
  - 0
  - 10
  - 20
  - 30
  - 40
  - 50

**Figure 4.7:** The results from enhancement network 3. Epochs are presented on the x-axis and the SNR and CNR values are presented on the y-axis. Every 5th epoch is included from 1 to 80 (including 80).

Figures 4.8 and 4.9 include a selection of results from enhancement network 3.
Figure 4.8: Test images from enhancement network 3. Original images are displayed to the left and their corresponding processed versions are on the right.
Figure 4.9: Test images from enhancement network 3. Original images are displayed to the left and their corresponding processed versions are on the right.
4.1.1 Histogram Equalization

HE and CLAHE were performed on the test set. The resulting images were compared against the corresponding original images to measure SNR and CNR difference in the same manner as described in section 3.4.1. These values were then compared against the results from the different versions of the enhancement network. The average SNR and CNR percentage differences were calculated for each technique. They are given in table 4.1.

<table>
<thead>
<tr>
<th></th>
<th>HE SNR difference</th>
<th>CLAHE SNR difference</th>
<th>HE CNR difference</th>
<th>CLAHE CNR difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-27%</td>
<td>-5.8%</td>
<td>+74%</td>
<td>-54%</td>
</tr>
</tbody>
</table>

In figure 4.10 the SNR differences are given for the various methods. The highest values of the different implementations of the enhancement networks are represented.

![Figure 4.10: The SNR percentage difference for the various methods](image)

In figures 4.11 the CNR values are found for the various methods. The highest values of the different implementations of the enhancement networks are represented.

![Figure 4.11: The CNR percentage difference for the various methods](image)
In Figure 4.12, a sample of the test images processed with HE can be seen.

Figure 4.12: Test images processed with HE. Original images are displayed to the left and their corresponding processed versions are on the right.
In figure 4.13, a sample of the test images processed with CLAHE can be seen.

Figure 4.13: Test images processed with CLAHE. Original images are displayed to the left and their corresponding processed versions are on the right.
4.1.2 Artifacts

No quantitative metrics were used to evaluate the results from the artifact reduction network. Instead, visual inspection was performed with the qualitative test. In figures 4.14 and 4.15 a selection of images that were processed by the artifact reduction network can be seen. These images, from the test set of A-images, were included in the qualitative test.

Figure 4.14: Test images from the artifact reduction network. Original images are displayed on the left and their corresponding processed versions are on the right.
Figure 4.15: Test images from the artifact reduction network. Original images are displayed on the left and their corresponding processed versions are on the right.
4.2 Qualitative Test

4.2.1 Test 1

In the first section of the test, the evaluators were tasked with classifying original images from processed images. 30 images were given to the evaluators, 15 original and 15 processed. The results are presented in tables 4.2 and 4.3. Table 4.2 displays the three medical students with less experience from CT images, and 4.3 represents the more experienced clinicians. The total average was an accuracy of 55.8%.

<table>
<thead>
<tr>
<th>Evaluator</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluator 1</td>
<td>11/30 ≈ 36.7%</td>
</tr>
<tr>
<td>Evaluator 2</td>
<td>13/30 ≈ 43.3%</td>
</tr>
<tr>
<td>Evaluator 3</td>
<td>14/30 ≈ 46.7%</td>
</tr>
<tr>
<td>Average</td>
<td>12.7/30 ≈ 42.2%</td>
</tr>
</tbody>
</table>

Table 4.3: Results from the experienced evaluators from qualitative test part 1.

<table>
<thead>
<tr>
<th>Evaluator</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluator 4</td>
<td>26/30 ≈ 86.7%</td>
</tr>
<tr>
<td>Evaluator 5</td>
<td>18/30 ≈ 60.0%</td>
</tr>
<tr>
<td>Evaluator 6</td>
<td>25/30 ≈ 83.3%</td>
</tr>
<tr>
<td>Evaluator 7</td>
<td>15/30 ≈ 50.0%</td>
</tr>
<tr>
<td>Evaluator 8</td>
<td>12/30 ≈ 40.0%</td>
</tr>
<tr>
<td>Average</td>
<td>19.2/30 ≈ 64.0%</td>
</tr>
</tbody>
</table>

4.2.2 Test 2

In the second part, the evaluators graded processed images by comparing them to the corresponding original image. They were graded on three separate scales: contrast, noise and artifacts. The criteria grades are presented in table 4.4. There were 28 R-image pairs and 14 A-image pairs. Contrast and noise grades were given to the R-image pairs and artifact and contrast grades were given to the A-image pairs. Figures 4.16, 4.17, 4.18 and 4.19 show histograms representing the results. One histogram corresponds to the less experienced medical students, and one to the experienced radiologists. The average results were calculated for each criteria. The coefficient of variation was calculated for each image with given grades.

<table>
<thead>
<tr>
<th>Grades</th>
<th>Contrast</th>
<th>Noise</th>
<th>Artifacts</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Significantly less contrast than original</td>
<td>Significantly more noise than original</td>
<td>Significantly more and/or larger artifacts</td>
</tr>
<tr>
<td>2</td>
<td>Slightly less contrast than original</td>
<td>Slightly more noise than original</td>
<td>Slightly more and/or larger artifacts</td>
</tr>
<tr>
<td>3</td>
<td>Identical or nearly identical to original</td>
<td>Identical or nearly identical to original</td>
<td>Identical or nearly identical to original</td>
</tr>
<tr>
<td>4</td>
<td>Slightly more contrast than original</td>
<td>Slightly less noise than original</td>
<td>Slightly less and/or smaller artifacts</td>
</tr>
<tr>
<td>5</td>
<td>Significantly more contrast than original</td>
<td>Significantly less noise than original</td>
<td>Significantly less and/or smaller artifacts</td>
</tr>
</tbody>
</table>
from the evaluators. The coefficient of variation varied from 0% to 46% for the test images. The average coefficient of variation was calculated for each criteria.

Figure 4.16 shows the contrast grades for the R-images. The total average score for all images and evaluators was 3.39. The average coefficient of variation was 25%.

Figure 4.16: Histograms representing the results from test 2: contrast. The medical students are represented to the left and the more experienced radiologists to the right.

Figure 4.17 visualizes the noise grades for the R-images. The total average score for all images and evaluators was 3.07. The average coefficient of variation was calculated to 25%.

Figure 4.17: Histograms representing the results from test 2: noise. The medical students are represented to the left and the more experienced radiologists to the right.

Figure 4.18 visualizes the artifact grades for the A-images. The total average score for all images and evaluators was 3.32. The average coefficient of variation was calculated to 31%.
4.2. Qualitative Test

Figure 4.18: Histograms representing the results from test 2: artifacts. The medical students are represented to the left and the more experienced radiologists to the right.

Contrast grades were also given to the A-images. These results are represented in figure 4.19. The total average score for all images and evaluators was 4.45. The average coefficient of variation was calculated to 13%.

Figure 4.19: Histograms representing the results from test 2: contrast in artifact images. The medical students are represented to the left and the more experienced radiologists to the right.

4.2.3 Test 3

In the third and last section of the qualitative test, the evaluators were given an original image and its corresponding processed image, without the information of which was which. Their assignment was to say which one they would prefer to receive from a CT scan to give a diagnosis. 40 image pairs were given, solely R-images. Table 4.5 displays the three medical students with less experience from CT images, and 4.6 represents the more experienced clinicians. The total percentage of processed images that were chosen was 48.1%.
Table 4.5: Results from qualitative test part 3. Visualizes how many processed images that were diagnostically preferred of the total 40 by the medical students.

<table>
<thead>
<tr>
<th>Evaluators</th>
<th>Originals</th>
<th>Processed</th>
<th>Processed percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluator 1</td>
<td>30</td>
<td>10</td>
<td>25.0%</td>
</tr>
<tr>
<td>Evaluator 2</td>
<td>9</td>
<td>31</td>
<td>77.5%</td>
</tr>
<tr>
<td>Evaluator 3</td>
<td>14</td>
<td>26</td>
<td>65.0%</td>
</tr>
<tr>
<td>Average</td>
<td>17.7</td>
<td>22.3</td>
<td>55.8%</td>
</tr>
</tbody>
</table>

Table 4.6: Results from qualitative test part 3. Visualizes how many processed images that were diagnostically preferred of the total 40 by the experienced clinicians.

<table>
<thead>
<tr>
<th>Evaluators</th>
<th>Originals</th>
<th>Processed</th>
<th>Processed percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluator 4</td>
<td>17</td>
<td>23</td>
<td>57.5%</td>
</tr>
<tr>
<td>Evaluator 5</td>
<td>29</td>
<td>11</td>
<td>27.5%</td>
</tr>
<tr>
<td>Evaluator 6</td>
<td>24</td>
<td>16</td>
<td>40.0%</td>
</tr>
<tr>
<td>Evaluator 7</td>
<td>10</td>
<td>30</td>
<td>75.0%</td>
</tr>
<tr>
<td>Evaluator 8</td>
<td>33</td>
<td>7</td>
<td>17.5%</td>
</tr>
<tr>
<td>Average</td>
<td>22.6</td>
<td>17.4</td>
<td>43.5%</td>
</tr>
</tbody>
</table>
Chapter 5

Discussion

This chapter discusses the results and the method of the project. In addition, it also includes a discussion of the feasibility of the proposed method in routine clinic. This is followed by the disadvantages and problems that have to be solved in that case. Future developments are then explored along with the aspects that could be improved given more time. It is also discussed how the methods can be altered to achieve this.

5.1 Discussion of Results

5.1.1 Enhancement Network

By looking at the SNR and CNR plots in section 4.1 it is difficult to immediately assess if one implementation of the network was better than the others. Looking at the training, validation and testing results together, it is confirmed that the results are inconclusive. It seems that the results from calculating the metrics have a variation that is too large, and thus no conclusion can be made on which method is superior. The resulting metrics from all plots move up and down seemingly at random throughout the training. From the plots of enhancement network 2 (figure 4.4) and enhancement network 3 (figure 4.7), the CNR trends appear to continue upwards, and that further training could keep improving the results. This is not certain, as the variation of the values is so large. The network is simply trying to translate the images from domain A to domain B, and the chosen metrics are not adequately connected to the success of this. The aim of enhancement network 3 was to add an additional objective for the network. It was supposed to keep improving its translations between the domain, and additionally increase the SNR and CNR of the output image. It is apparent from figure 4.7 that this did not succeed, as the SNR and CNR keep varying from epoch to epoch.

A conclusion that can be made is that the overall output from all networks is positive. An increase of both SNR and CNR can be seen for the majority of the later epochs from all training sessions. At most, the networks output an overall increase in CNR of approximately 60%, 40% and 40% each, as seen in figure 4.7, 4.4 and 4.7 respectively. An improvement is occurring based on the metrics.

When analyzing the images from the different implementations of the enhancement network visually (figures 4.2, 4.3, 4.5, 4.6, 4.8 and 4.9), much can be said. The difference between the images of the different network implementations is small. The contrast seems to increase in a majority of the images. The effect is greatest when the initial contrast is low. For high contrast images, as expected, there is not much improvement. The output contrast level seems to be quite consistent between images, which is a positive aspect as the 2D slices must be put back together to form the original 3D image.

In some of the images there are slight structural changes of the anatomy within the image.
The exact reason for why this happens is difficult to say, but it is a clear disadvantage of this method. Some structures are added and some removed, both cases are unfavorable. What can be seen, however, is that the morphology of the cardiac ventricles remain intact for all images, it is never removed. Small structures are sometimes added to this region, which are often quite simple to detect. However, this could still pose problems when utilizing the method.

The noise levels seem to be slightly reduced. However, the noise reduction can have a negative effect. It can blur out edges and patterns, which could remove important details in the images. This was also noticed during the third part of the qualitative test, were many original images were preferred because of this effect in the corresponding processed images. A possible solution in future development could be to use domain B images with the same noise structure as the domain A images, or add some noise to the domain B images. The goal would be to have the same or similar noise levels in the two domains, so that the network does not manipulate it and only focuses on the contrast. Alternatively, the tasks of the enhancement network could be given to two separate networks. One network would then focus on contrast enhancement while the other focused on noise reduction. This division is discussed further in section 5.3.

Histogram Equalization

Comparing the performance of HE and CLAHE with the enhancement network on the test set yielded many interesting results (as seen in section 4.1.1), some of which could be expected. By visual comparison there can be no doubt that the CycleGAN yields superior output. The majority of the structure from the original image is maintained with the enhancement network, with the main difference that the output image has increased contrast in the cardiac ventricles. The noise is increased notably when using HE (a decrease in SNR by 27%), making it more difficult to distinguish borders such as the myocardium, which is a severe negative effect. CLAHE has a decrease in SNR by 5.8%, and in many of the images the noise amplification is quite apparent. The CycleGAN maintains the SNR difference positive throughout most of the trainings for the different network implementations. A disadvantage of the CycleGAN is the possibility of structural changes, while no such transformation should be possible with HE and CLAHE.

Even though the overall CNR increase is higher with HE, this is not necessarily a fair portrayal of image quality. The purpose of HE is contrast increase, but it is a hard-coded algorithm performed without previous knowledge of cardiac CT images. From visual inspection of images in the test set processed with HE (figure 4.12), it is clear that the brightness of the heart is increased, but not necessarily the contrast of the ventricles, which is what is actually desired. This suggests that, as predicted, the CNR metric is suboptimal. It does not evaluate the ventricular contrast in the desired manner, as the CNR calculations only use the contrast between the heart and the background, and not within the heart. Since HE increases the brightness of heart so considerably, it acquires the highest CNR increase, even though it does not acquire high ventricular contrast. CNR was however the best available quantitative evaluation of the contrast. An automatic segmentation of the heart was unavailable, but would have been preferable. The information of which pixels that represent the ventricles and which represent the myocardium would then have been attainable.

From visual inspection (figure 4.13), the test images processed with CLAHE does not seem to have increased ventricular contrast, in fact, it often seems to decrease. CLAHE has a CNR decrease of 54% on the test set which suggests the same. The brightness and overall visibility in these images are clearly decreased. What is consistent with the CLAHE images is that the septum is darkened which slightly increases the ventricular contrast. The CycleGAN ap-
5.1. Discussion of Results

pears more capable of producing images with a higher diagnostic quality, being an intelligent network in the sense that it can handle many different input images in a custom way.

5.1.2 Artifact Reduction Network

When analyzing the images from the artifact reduction network (figures 4.14 and 4.15), it is evident that many artifacts are reduced, some almost completely. Some are, however, only smeared out or completely intact. The contrast and brightness in the output images are often highly increased from the input image. In these images, the artifact seems to disappear within the increased brightness. As the artifacts in almost all of these images has a very high pixel intensity compared to the remainder of the image, it seems natural that it blends in and disappears to a higher degree when the rest of the image has increased pixel intensities. The training images in the B-domain were high quality R-images, most with high contrast. The network thus tried to achieve both of these aims.

What we do see in the artifact reduction results is some considerable structural change within the images, to a higher degree than the implementations of the enhancement network. If the results from this project were to be used in the industry, it is of high importance that the clinicians are aware of this. A simple solution that could be used for this problem would be to output the original image alongside the processed image, which could be used for comparison. If the structure is unaltered between the images, the clinician can be certain that no structural change has occurred.

5.1.3 Qualitative Test

Test 1

The average accuracy was 55.8%, a result that suggests that there are considerable similarities between the original and the processed images. The lower numbers were mostly from the more inexperienced evaluators who had no significant training with CT. These evaluators had an average accuracy of 42.2%. The more experienced evaluators scored an accuracy of 64%, a considerable difference. However, lower results were also obtained by the more experienced clinicians. These results suggest that the processed images are quite similar to the originals, but in many cases contain noticeable differences and cannot pass as unaltered CT images to an experienced radiologist. When questioning one of the experienced evaluators with high accuracy, it became clear that the main difference that signified the image was the noise pattern. When the noise was smoothed out in the processed images, the images became less sharp. This is quite apparent and reduces visibility of image details such as edges.

Test 2

For every category evaluated, the grade 5 was the highest score and 1 the lowest. The contrast received an average grade of 3.39, which corresponds to slight improvement of the contrast, and confirms from the CNR measurements that the implemented network truly achieves a contrast increase. The noise reduction received an average grade of 3.07, which corresponds to the overall noise levels remaining unchanged. The artifact reduction received an average grade of 3.32. These results show a slight overall artifact reduction. The total contrast average for the A-images was 4.45, representing a contrast increase between slight and considerable magnitude. This confirms from the qualitative test that the contrast is clearly increased by the artifact reduction network.

The results had a high variation between the evaluators. Not only in overall score, but the results varied to a high degree between separate images. This shows once again how subjective image quality is, and that the results cannot rely on visual analysis alone. The average coeffi-
5.2. Discussion of Method

5.2.1 CycleGAN and Input Data

Since the only information the network was given to work with and learn from was the input data, the definitions of domain A and domain B were highly important. Two initial approaches were performed and compared. One method was to exclude the middle portion containing images that did not really belong to either of the two classes, enhancement network 1, and the other was to use the entire data set of images, enhancement network 2. The positive aspect of using only the top and bottom portion of the data set was that the network was assumed to more distinctly learn the difference between the two domains. If images are included in both classes that does not truly belong there, the network could learn incorrectly from these images. With the high number of images that were used in this project, there was bound to be a significant amount of overlapping. Images with higher quality would be classified to the low quality class and vice versa. The negative aspect of excluding the middle portion was the problem of how the network would handle these types of images (medium quality images) after training. Since it never would have seen any similar images during training, it might not know how to process them. A theory was that if the network could handle images with very poor quality, then it should also be able to handle images that were mediocre. In the end both approaches were implemented, and based on the validation set (image quality metrics and visual inspection) the most successful one was deemed to be the strategy of including the entire data set. When the middle portion was excluded, the output images were to a higher degree deformed, leading to the conclusion that the network had overfitted because of lack of generalization in the input. The next implementation of the network was built on the network that trained on the entire data set, enhancement network 2.

The next approach was an adjusted cost function, enhancement network 3. The acquired difference of SNR and CNR from the original image was added to the cost function, so that the cost was increased when the SNR and CNR difference was low. No significant improvement of the results was obtained. In the end, the network without the adjusted cost function, enhancement network 2, was chosen to perform the qualitative test.
Another discussion point throughout the project was whether all the 2D slices should be given to the network, even the ones that did not display the cardiac ventricles. Many of the 2D-images added to the training sets only showed the apex and other regions of less interest. It is not possible to assess if such an image has any contrast between the ventricles. By excluding it, perhaps the network could focus its learning on the areas of interest, images containing the cardiac ventricles. Using a smaller number of images for the training would additionally reduce the training time. The decision was, however, to include all the slices from the 3D-image in the training, validation and test data, to let the network learn how to handle all images. An option could be to implement an algorithm that finds and removes the images that are not of interest, and only sends the most interesting images to the system. This was not implemented in this project, but could perhaps be implemented in future development.

Separately Scaled Data

Throughout the project it was assumed that the contrast of the image slices within the same time frame was constant. For this reason, only one middle slice of each time frame was analyzed before classifying the entire time frame to a class. If the middle slice of the time frame had high enough contrast, the entire 3D image of approximately 100-400 slices was classified to the high contrast domain. Towards the end of the project, it was discovered that this assumption was incorrect, as some input images within the same time frame had a considerable variance in contrast.

Additionally, the images that had been received for the project had not been adjusted to the Hounsfield scale or a general scale common to all images. Instead, each image had been given an intensity scale that was set individually. The scale was set from 0 to 255, and each image had its highest intensity point set to 255. By setting each image to a separate scale, the rest of the pixel intensities of the image were affected by the intensity of the maximum intensity pixel. If the image had a very bright point as the maximum intensity, e.g., the images with artifacts, the rest of the pixel values were lowered to fit to the scale and to adjust to a very high intensity as a maximum value of 255. This meant that the brightness and contrast of these images were considerably lowered in the regions of interest. The primary problem that this caused was the errors in the classification, but also that the network did not train on the images as they should have been perceived. To what extent this might have had an impact on the output images is difficult to assess, but it can be assumed that the results from the artifacts network were the most influenced as those input images were affected the most. In future development, the images should be acquired with a general scale to minimize the risk of impact.

Image Quality Metrics

The metric used for evaluating the contrast enhancement was CNR. To calculate it, the ROI was defined as an area containing the heart. The ROI was then compared to the background of the image to calculate the CNR according to equation 2.6. This calculation is suboptimal as seen from the images processed with histogram equalization in figure 4.12. These images have extremely high brightness within the heart, but since all of the heart pixels are affected, the actual contrast within the heart is decreased and the quality is quite poor. The CNR only calculates the contrast between the ROI and the background, giving the HE images high values. Hence, it is clear that acquiring the highest CNR value does not correspond to acquiring the best ventricular contrast. The optimal quantitative evaluation could be a segmentation of the heart that could identify the ventricular and the myocardial pixels to compare the contrast between these instead. However, no such technique was available, and CNR was used instead. It worked as an adequate approximate evaluation, but was obviously flawed.
5.3 Future Development

The initial idea of the final project product was a pipeline with the networks where an input image were first processed by the artifact reduction network, and then the enhancement network. The image would receive artifact reduction, contrast increase and noise reduction by going through the pipeline. Optimally, the enhancement network would be divided into two separate networks: one that focused on noise reduction and one that focused on contrast enhancement. In the noise network the performance could be measured with peak signal to noise ratio (PSNR) and other image quality metrics that compare the similarity between two images. This was done in the project by Wolterink et al. [24], where the PSNR was measured in the generated images by comparing them to the corresponding routine-dose CT images as references. This can, however, be difficult without paired data. Generated images cannot be compared to corresponding routine-dose CT images if there are none. The solution would be to take original high quality images with low noise levels and to add noise to them. The original and the image with added noise could then be observed as a corresponding image pair with high and low noise levels, respectively. The images with added noise would pass through the implemented network for training and testing to produce generated high quality images. The PSNR could then be measured by comparing the generated image with the original high quality image as reference. The contrast network could receive training images with similar noise levels and patterns to both of its domains. The only difference between the domains would be the difference in ventricular contrast levels. The contrast network would then focus solely on contrast enhancement, while the noise network would focus solely on noise reduction. An illustration of the intended pipeline with the three networks can be seen in figure 5.1 [24].

![Figure 5.1: The plan for the final pipeline of the finished product. This was not included in the scope of the project.](image)

Continued work with the contrast network could increase the performance, with continued alterations of its hyperparameters and input images. The input images to the artifact reduction network can be altered in many different ways, as mentioned in 5.1.2. It would be interesting to analyze the results when exchanging the high quality images in the artifact reduction network for images without artifacts but with low contrast. Perhaps the network would then solely focus on removing and reducing the artifacts instead of also increasing the
contrast of the image. It is then of course quite possible that the results would be inferior to the current results. Another strategy would be to use images with high contrast and low noise levels in both image domains, with the only difference of artifacts occurring in the low quality domain. Perhaps the network would be more focused on removing artifacts instead of altering other features.

The results would probably improve by using the 3D-images as input to the networks instead of 2D. Working with the entire CT image instead of a few slices would give the network a much more wholesome picture of the situation. The output would also be much more homogeneous, as the entire 3D-image would be processed in the same manner, instead of processing one 2D slice at a time.

Another observation is that the networks might perform better if the outside borders of the CT images were removed. This outside area is a remainder from the CT scan and does not contain any anatomical information. This gives information to the network of the appearance of the images that is completely redundant, and might alter the results in unforeseeable ways.

5.4 Discussion of the Research Questions

The answers to the research questions are given here, to conclude whether the aims were achieved.

Can an efficient deep learning method be used to enhance the contrast and reduce noise levels in dose modulated CT images?

The short answer is yes. The measured metrics for the different versions of the enhancement network show an overall increase of CNR throughout the training sessions. At most, the different implementations of the enhancement network output an overall increase of approximately 60%, 40% and 40% each. By visual inspection the ventricular contrast is clearly increased in many of the R-images and A-images, and remains constant in others. Only a few actually lose contrast. The results from the qualitative test suggests a slight overall increase of contrast from enhancement network 2, and a considerable increase from the artifact reduction network.

The measured metrics show an overall slight increase of SNR from all versions of the enhancement network. At most, the networks achieve an SNR increase of approximately 10%. By visual inspection, it can be seen that noise is reduced in many of the images, but not in all. The results from the qualitative test suggest that the overall noise levels remain constant, but the results are varying from image to image, and the coefficient of variation was calculated to 25%.

Can a similar method be implemented for mitigating artifacts in dose modulated CT images?

The short answer is yes. It is visible from some of the output images that artifact appearance was mitigated, and in some cases essentially removed, as the artifact disappears into the increased pixel intensities of the output image. However, the results are mixed and many artifacts remain intact after processing. The average results from the qualitative test are slightly above 3, suggesting a slight overall increase. Many images received quite high marks while many received quite low, and the coefficient of variation was calculated to 31%. The conclusion is that the network performs well for certain images, but not all.
Are the enhanced images suitable for diagnosis?

The short answer is no, not yet. Due to the overall problem with structural alterations and the unreliability of the results, the output images are not fit for diagnosis and the implemented networks are not ready for medical use. However, the results look promising after only this short project, and continued development with this implementation could possibly yield reliable output. By manipulating the image domains of the networks much can be achieved, and the results could likely be improved further.
Chapter 6

Conclusion

The purpose of this thesis was to use a deep learning method to improve cardiac CT image quality. The aspects on which to focus were contrast, noise and artifacts. The results show that the implemented CycleGAN networks increases ventricular CNR and SNR to some extent, and artifacts are mitigated adequately for many images. However, from visual inspection and the qualitative test it is clear that the image quality enhancement is not consequent for all images. Issues with the technique were also observed, such as anatomical structure alterations. The method shows promise but needs continued development before it can be applied to medical usage.
Bibliography


