Low-dose contrast CT reconstruction with side information



Rita Abreu Romani
Mestrado Ciência de Computadores,
Departamento Ciência de Computadores
Faculdade de Ciências da Universidade do Porto
2023/2024

U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

Low-dose contrast CT reconstruction with side information

Rita Abreu Romani

Dissertação realizada no âmbito do Mestrado Ciência de Computadores,
Departamento Ciência de Computadores,
2023/2024

Orientador

Francesco Renna, Professor Auxiliar, Faculdade de Ciências da Universidade do Porto

Coorientador

João Manuel Pedrosa, Professor Auxiliar Convidado, Faculdade de Engenharia da Universidade do Porto

Abstract

Coronary Artery Disease (CAD) occurs when the coronary artery becomes clogged with plaque, restricting the blood flow to get to the heart muscles. Genetic or familial factors can be worsened by high cholesterol and diabetes influenced by this condition. CAD it is possible to be detected using Computed Tomography (CT), what allows the evaluation of the quantity of calcium from CT. This technique provides an unobtrusive view of the internal structures. While calculating the quantity of calcium in the coronary arteries has prognostic value, it is not used for diagnosis. The diagnosis of CAD requires a CT scan with contrast to measure stenoses. The amount of coronary artery stenosis can be determined by injecting intravenous contrast agents. However, the use of contrast-enhanced CT incurs considerable hospital expenses and exposes patients to large radiation doses. Fortunately, through reconstruction methods, it is possible to reduce the radiation dose used for contrast-enhance CT.

This dissertation investigates deep learning (DL) solutions for low-dose CT reconstruction in the presence of side information, namely, CT without contrast, which is usually collected previous to contrast-enhanced CT.

The main novelty of this dissertation resides in the exploration of DL solutions for low-dose CT reconstruction that are able to take advantage of the presence of available side information by incorporating it into the U-Net architecture.

Resumo

A CAD ocorre quando a artéria coronária fica obstruída com placas, restringindo o fluxo sanguíneo para chegar aos músculos do coração. Os fatores genéticos ou familiares podem ser agravados pelo colesterol elevado e pela diabetes influenciada por esta condição. CAD é possível ser detectado através da CT, o que permite avaliar a quantidade de cálcio do CT. Esta técnica fornece uma visão discreta das estruturas internas. Embora o cálculo da quantidade de cálcio nas artérias coronárias tenha valor prognóstico, não é utilizado para diagnóstico. O diagnóstico de CAD requer uma CT com contraste para medir as estenoses. A quantidade de estenose da artéria coronária pode ser determinada através da injeção de agentes de contraste intravenosos. No entanto, a utilização de CT com contraste incorre em despesas hospitalares consideráveis e expõe os doentes a grandes doses de radiação. Felizmente, através de métodos de reconstrução, é possível reduzir a dose de radiação utilizada para CT com realce de contraste.

Esta dissertação investiga soluções DL para a reconstrução de CT a baixas doses na presença de informação secundária, i.e., CT sem contraste, que é geralmente recolhida antes do CT com contraste .

A principal novidade desta dissertação reside na exploração de soluções DL para reconstrução de CT a baixas doses que sejam capazes de tirar partido da presença de informação lateral disponível, incorporando-a na arquitetura U-Net.

Acknowledgments

I would like to express my gratitude to my supervisors, Professor Francesco Renna and Professor João Pedrosa, for their invaluable assistance. Without their guidance, this dissertation would not have been feasible. I would want to express my gratitude to my parents and my friends for their support over the duration of this endeavor ...

Contents

Abstract i

Resumo iii

Acknowledgments v

Contents viii

List of Tables ix

List of Figures xiv

Listings xv

Acronyms xvii

1 Introduction 1

1.1 Main goals 2

1.2 Dissertation layout 3

2 Background 5

2.1 Coronary Artery Disease 5

2.2 Deep Learning (DL) 6

2.2.1 Artificial Neural Networks (ANN) 7

2.2.2 Training an ANN: Process and Techniques 9

2.2.3 Convolutional Neural Networks (CNN) 12

2.2.4 U-Net . . . 14
2.3 Image Reconstruction in CT Scanning . . . 16
2.3.1 Image Acquisition and Reconstruction Process . . . 16
2.3.2 Filtered Back Projection (FBP) . . . 17
2.3.3 DL for Image Enhancement . . . 18

3 Literature Review **19**
3.1 Discussion . . . 28

4 Methods **31**
4.1 Materials . . . 31
4.2 Data pre-processing . . . 32
4.3 Architectures and Training . . . 33
4.3.1 Computation of CT Measurements and Filtered Backprojection . . . 34
4.3.2 Reconstruction without side information . . . 34
4.3.3 Reconstruction with side information using 2 channels . . . 35
4.3.4 Reconstruction with side information using 2 encoders . . . 36

5 Results and Analysis **39**
5.1 Results of the U-Net without side information . . . 40
5.2 Results of the U-Net with two input channels . . . 48
5.3 Results of the U-Net with two encoders . . . 56
5.4 Comparing image with the 3 different models . . . 64
5.5 Discussion . . . 64

6 Conclusion **67**

Bibliography **69**

List of Tables

3.1 Summary of Papers on Image Denoising Techniques 30

4.1 Orca Data. 32

4.2 Summary of the Orca Dataset's Paired Filtration. 32

5.1 Comparison of MSE and SSIM metrics for the U-Net model under different configurations (256, 128, and 64 lines). The table includes minimum, maximum, mean, and standard deviation values for both test and validation datasets. 40

5.2 Comparison of MSE and SSIM metrics for the U-Net model under different configurations (256, 128, and 64 lines). The table includes minimum, maximum, mean, and standard deviation values for both test and validation datasets. 48

5.3 Comparison of MSE and SSIM metrics for the U-Net model under different configurations (256, 128, and 64 lines). The table includes minimum, maximum, mean, and standard deviation values for both test and validation datasets. 56

List of Figures

2.1 Coronary arteries of the Heart [14]. 5

2.2 Showing the formation of cholesterol and plaque at a rip in the artery's inner lining, illustration of arterial atherosclerosis development This accumulation may harden, constrict the artery lumen, lower blood flow, and maybe cause a total blockage by a clot [15]. 6

2.3 A typical DL workflow to solve real-world problems [22]. 7

2.4 ANN structure diagram [10]. 8

2.5 ReLu example [10]. 8

2.6 Leaky ReLu example [10]. 8

2.7 Tanh example. [10] 9

2.8 Sigmoid example. [10] 9

2.9 Batch Normalization [16]. 12

2.10 CNN [16]. 12

2.11 Illustration of the convolution operation between the input feature map and the kernel in a CNN convolutional layer [18]. 13

2.12 U-Net is an encoder-decoder segmentation network with skip connections [20]. . . 14

2.13 U-Net architecture [20]. 15

2.14 DL-based image reconstruction and post-processing methods in positron emission tomography for low-dose imaging and resolution enhancement [3]. 16

2.15 FBP. The original image (left) and a CT scan after dose reduction, display increased noise and artifacts(right) [23]. 18

3.1 Architecture of the network proposed in [7]. 20

3.2 Architecture overview of MAGUNet, comprising a multi-scale feature extraction block (MFE), attention-guided filter block (AGB), and residual denoising block (RDB) [13]. 21

3.3 Architecture overview of Residual U-Net [9]. The network leverages multi-scale feature extraction to capture detailed image priors across different resolution levels. 22

3.4 Overall framework of the proposed DU-GAN[26]. 22

3.5 Overview of SAGAN [5]. 23

3.6 Proposed generator of the SAGAN. The residual block in the center of the network is repeated K times, and K was chosen as 9 for the experiment [25]. 24

3.7 SAGAN [5]. 25

3.8 The process for wavelet-assisted noise disentanglement by extracting high-frequency sub-band images from a CT image using k-level wavelet transform and inverse wavelet transform [5]. 26

3.9 Proposed deep CNN architecture for wavelet domain denoising [11]. 27

4.1 schemes of the reconstruction without side information [27]. 35

4.2 schemes of the reconstruction with side information using 2 channels [27]. 36

4.3 schemes of the reconstruction with side information using 2 encoders [27]. 36

5.1 U-NET Images with the 3 best (left) and the 3 worst (right) MSEs (256 lines, Test Set). 42

5.2 Corresponding FBP Images related to the ones above (256 lines, Test Set). . . . 42

5.3 U-NET Images with the 3 best (left) and the 3 worst (right) SSIMs (256 lines, Test Set). 43

5.4 Corresponding FBP Images related to the ones above (256 lines, Test Set). . . . 43

5.5 U-NET Images with the 3 best (left) and the 3 worst (right) MSEs (128 lines, Test Set). 44

5.6 Corresponding FBP Images related to the ones above (128 lines, Test Set). . . . 44

5.7 U-NET Images with the 3 best (left) and the 3 worst (right) SSIMs (128 lines, Test Set). 45

5.8 Corresponding FBP Images related to the ones above (128 lines, Test Set). . . . 45

5.9 U-NET Images with the 3 best (left) and the 3 worst (right) MSEs (64 lines, Test Set). 46

5.10 Corresponding FBP Images related to the ones above (64 lines, Test Set). 46

5.11 U-NET Images with the 3 best (left) and the 3 worst (right) SSIMs (64 lines, Test Set). 47

5.12 Corresponding FBP Images related to the ones above (64 lines, Test Set). 47

5.13 U-NET Images with the 3 best (left) and the 3 worst (right) MSEs (256 lines, Test Set). 50

5.14 FBP Images corresponding to the U-NET images above (256 lines, Test Set). . . 50

5.15 U-NET Images with the 3 best (left) and the 3 worst (right) SSIMs (256 lines, Test Set). 51

5.16 FBP Images corresponding to the U-NET images above (256 lines, Test Set). . . 51

5.17 U-NET Images with the 3 best (left) and the 3 worst (right) MSEs (128 lines, Test Set). 52

5.18 FBP Images corresponding to the U-NET images above (128 lines, Test Set). . . 52

5.19 U-NET Images with the 3 best (left) and the 3 worst (right) SSIMs (128 lines, Test Set). 53

5.20 FBP Images corresponding to the U-NET images above (128 lines, Test Set). . . 53

5.21 U-NET Images with the 3 best (left) and the 3 worst (right) MSEs (64 lines, Test Set). 54

5.22 FBP Images corresponding to the U-NET images above (64 lines, Test Set). . . . 54

5.23 U-NET Images with the 3 best (left) and the 3 worst (right) SSIMs (64 lines, Test Set). 55

5.24 FBP Images corresponding to the U-NET images above (64 lines, Test Set). . . . 55

5.25 U-NET Images with the 3 best (left) and the 3 worst (right) MSEs (256 lines, Test Set). 58

5.26 FBP Images related with the image above. 58

5.27 U-NET Images with the 3 best (left) and the 3 worst (right) SSIMs (256 lines, Test Set). 59

5.28 FBP Images related with the image above. 59

5.29 U-NET Images with the 3 best (left) and the 3 worst (right) MSEs (128 lines, Test Set). 60

5.30 FBP Images related with the image above. 60

5.31 U-NET Images with the 3 best (left) and the 3 worst (right) SSIMs (128 lines, Test Set). 61

5.32 FBP Images related with the image above. 61

5.33 U-NET Images with the 3 best (left) and the 3 worst (right) MSEs (64 lines, Test Set). 62

5.34 FBP Images related with the image above. 62

5.35 U-NET Images with the 3 best (left) and the 3 worst (right) SSIMs (64 lines, Test Set). 63

5.36 FBP Images related with the image above. 63

5.37 image without side information (left), image with side information using 2 channels (center) and image with side information using 2 encoders (right) 64

Listings

Acronyms

DCC Departamento de Ciência de Computadores

FCUP Faculdade de Ciências da Universidade do Porto

CAD Coronary Artery Disease

CT Computed Tomography

DL Deep Learning

ANN Artificial Neural Networks

MSE Mean Squared Error

SSIM Similarity Index Measure

FBP Filtered Back Projection

CNN Convolutional Neural Network

GANs generative adversarial networks

PSNR Peak to Signal Noise

RDB Residual denoising block

SAGAN Attention Generative Adversarial Network

MFE Multi-scale Feature Extraction

AGB Attention-Guided Filter Block

Chapter 1

Introduction

Recognized as the main cause of death in developed countries, CAD results have estimated 7.8 million deaths [17]. The Centers for Disease Control and Prevention estimate that CAD is a serious health concern in the United States, causing mortality annually of 659,000 people [1]. Atherosclerosis is what causes this disease. This is a biological process in which cholesterol builds up behind the endothelial lining of blood vessels, causing plaques to form that can block blood flow through the arteries.

One of the techniques that provide the possibility of visualizing without damaging the heart vessels is Contrast-enhanced CT and CT, as main objective of these approaches is to provide a clear and detailed view of the coronary arteries without resorting to invasive procedures like coronary angiography, which involves inserting a tube into a blood artery [4]. Both techniques (Contrast-enhanced CT and CT) are very precise, and can therefore improve the prognosis of patients. They play a crucial role in the diagnosis and treatment of CAD.

A method widely used in medical imaging, which helps to improve the sharpness and visibility of the identification of anatomical structures, namely the heart, is Contrast enhancement. One of the procedures used to improve the detection of certain areas for imaging studies such as CTis often administered by injecting contrast agents into the body. These agents achieve a better result in the precision of the images, which allows in specific locations such as the blood lumen and the blood-thrombus interface, these agents work by increasing the absorption of CT rays. However, it is not necessary to inject contrast agents in all procedures for the purpose of contrast enhancement.

Another advantage of Contrast-enhanced CT is that it can identify the presence of coronary calcium, as it can provide important data regarding calcification of the coronary arteries, but cannot locate arterial blockages. One of the ways to improve tissue attenuation and more detailed images was developed to resolve this impediment, contrast-enhanced CT, which means that contrast agents will be used to increase tissue density.

Although contrast-enhanced CT can offer greater imaging capabilities, it has some disadvantages such as renal toxicity, renal failure, and other kidney diseases.

Low-dose CT is a technology meant to minimise the patient's radiation exposure during imaging operations, without thus affecting the dosage of contrast agent given. This effort mainly aims to provide techniques for reduced radiation dosage reconstruction of high-quality contrast-enhanced CT images. Within this framework, the idea of a sinogram is crucial. During a CT scan, a sinograma graphical depiction of the raw data acquired from many angles is subsequently utilized to rebuild the final image. Since each point in the sinogram corresponds to a projection obtained with a low radiation dosage, in low-dose CT the image reconstruction procedure uses less data points in the sinogram. The fundamental goal of our work is to sustain image quality and diagnostic accuracy in spite of this decrease.

Utilizing DL algorithms to reconstruct low-dose CT images is an effective approach to achieving this reduction in radiation. This method potentially reduces the patient's radiation exposure by reducing the radiation dosage through the use of supplementary data. This investigation investigates the feasibility and efficacy of employing DL techniques to improve low-dose CT imaging. The goal is to create a diagnostic tool that is more efficient and reliable in identifying CAD.

1.1 Main goals

The study's goal is to find ways to deal with the problems that come with low-dose CT imaging, specifically by reducing unwanted noise and other artefacts that might make it hard to see important medical information. This could be done in part by adding more useful information. When more data is added to the rebuilding process, the quality and accuracy of the images that are made improve. CT scans help people understand what they see better and can be used for more things. This helps doctors come up with more accurate diagnoses and choose better ways to treat patients.

Advanced computer algorithms and mathematical models have been developed throughout the process to make it easy to add new data to the practice of CT scan reconstruction. These approaches carefully combine low-dose CT data with additional information about the scene in an effort to limit the effect noise and errors that are inherent in the data.

This work will aim to how utilizing additional available information can substantially improve the effectiveness of CT reconstruction from low-dose data. This method aims to improve the quality and value of reconstructed images, this is an opportunity to improve the field of medical imaging and consequently improve patient outcomes, all of which can be achieved through more accurate and reliable diagnostic tools.

The objectives of the study presented in this dissertation are the following:

- **Improving CT Reconstruction with Low-Dose Data: Focus on Side Information Integration**

The objective of this investigation is to use CT how we can improve the quality of image reconstruction including low-dose data combined with other data such as contrast-free CT scan, in order to minimize radiation exposure and at the same time to preserve or improve the accuracy of reconstructed images for diagnosis.

- **Comparing U-Net Architectures: Single vs. Dual Input Approaches for Enhanced CT Reconstruction Performance**

  The objective of this investigation will be to compare different U-Net models for CT reconstruction, focusing especially on the way in which U-Net models behave with just one input and with double input that encompasses additional information. The ultimate goal is to discover the best approach to increase the quality of CT reconstruction.

1.2 Dissertation layout

This dissertation consists of six chapters, which are structured as follows:

- Chapter 2: Key concepts necessary to understand the work, encompassing the fundamentals of machine learning, DL, U-Net approaches, and background information on CAD and their detection.
- Chapter 3: Analysis of the recent research on the use of DL for low-dose CT reconstruction.
- Chapter 4: Description of the dataset that is used to run the U-Net algorithm and the preprocessing that was performed on them. This chapter explains the proposed reconstruction architectures and their essential components.
- Chapter 5: Presents the measures used to evaluate the performance of the proposed models. The results and subsequent analysis are presented and evaluated in detail.
- Chapter 6: Concludes with a summary of the findings and a discussion of potential avenues for further research and development.

Chapter 2

Background

This chapter gives a short medical description of CAD and a brief introduction to machine learning and DL. Prior to discussing DL, there is a section dedicated to CT image reconstruction. The significance lies in the use of DL techniques discussed in this work to enhance the process of reconstructing CT images.

2.1 Coronary Artery Disease

CAD is the main focus of this dissertation, and this disease is also known as ischemic heart disease. CAD is a medical illness defined by the buildup of plaque inside the coronary arteries, the blood channels that carry oxygen-rich blood to the heart muscle (see Figure 2.1). Plaque is composed of fat, cholesterol, calcium, and other blood-derived compounds. Plaque accumulation may constrict and stiffen the arteries over time, reducing blood flow to the heart. This may ultimately result in chest discomfort (angina), shortness of breath, and a heart attack. Worldwide, CAD is the primary cause of mortality and disability [19].

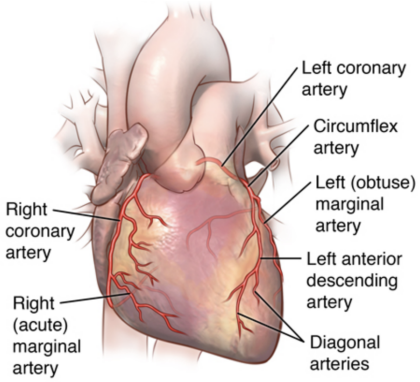


Figure 2.1: Coronary arteries of the Heart [14].

As illustrated in Figure 2.2, arterial atherosclerosis may occur when deposits of cholesterol and plaque accumulate at a tear in the inner lining of an artery. As the deposits harden and

occlude the arterial lumen, blood flow to distant tissues decreases, and a clot may become lodged, completely blocking the artery [15].

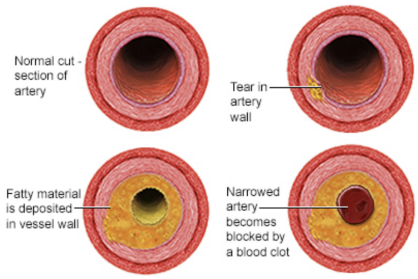


Figure 2.2: Showing the formation of cholesterol and plaque at a rip in the artery's inner lining, illustration of arterial atherosclerosis development This accumulation may harden, constrict the artery lumen, lower blood flow, and maybe cause a total blockage by a clot [15].

Often, a CT scan is utilized to diagnose coronary artery disease. This approach is a non-intrusive way to peek into structures. The concept behind this technique is to deliver X-rays or other forms of ionizing radiation across a region of interest at present. Depending on the tissue's thickness and density, the X-rays are attenuated during the same time they pass through it. When the attenuation is high, the CT gets brighter, and when it is low it gets darker.

Often referred to as calcium scoring or coronary artery calcium (CAC) scans, quantifying the calcium content in the coronary arteries using specialist CT scans is a well-established and extensively used test for CAD. These tests quantify calcified plaque in the arteries without the use of contrast chemicals. Calcium score is computed using CT image calcification density and area as basis. Given a more badly blocked artery, a higher calcium score signifies a higher risk of CAD. This approach has a drawback, however, as some people especially those younger than middle age may have CAD without any obvious abnormalities.

Image contrast is produced by contrast agents that increase the attenuation of nearby tissues. This necessitates the use of an intravenous contrast agent, which is used to augment the attenuation and contrast between tissues and the artery. Several people are allergic to iodine based contrast agents, hence the use of this contrast has certain disadvantages. The administration of this medication involves the insertion of a needle, which may cause pain and skin irritation/damage in the patient.

2.2 Deep Learning (DL)

DL is one of the most fascinating subfields in machine learning when ANN with many layers handle data from inputs. DL is the training of a multi-layer neural network aiming at data pattern detection.

Identifying patterns in data can be quite challenging for individuals, but DL algorithms excel at this task by effortlessly recognizing them. Among the various challenging tasks, such as

natural language processing, image and audio recognition, and more, they can achieve impressive levels of accuracy. Refer to the study done by Taye et al. [22] for a thorough understanding of the topic.

One of the major advantages of DL is its ability to independently learn features from raw data, without requiring human interaction or pre-designed characteristics. Transformers, recurrent neural networks (RNNs), and CNN are well recognized DL models [22].

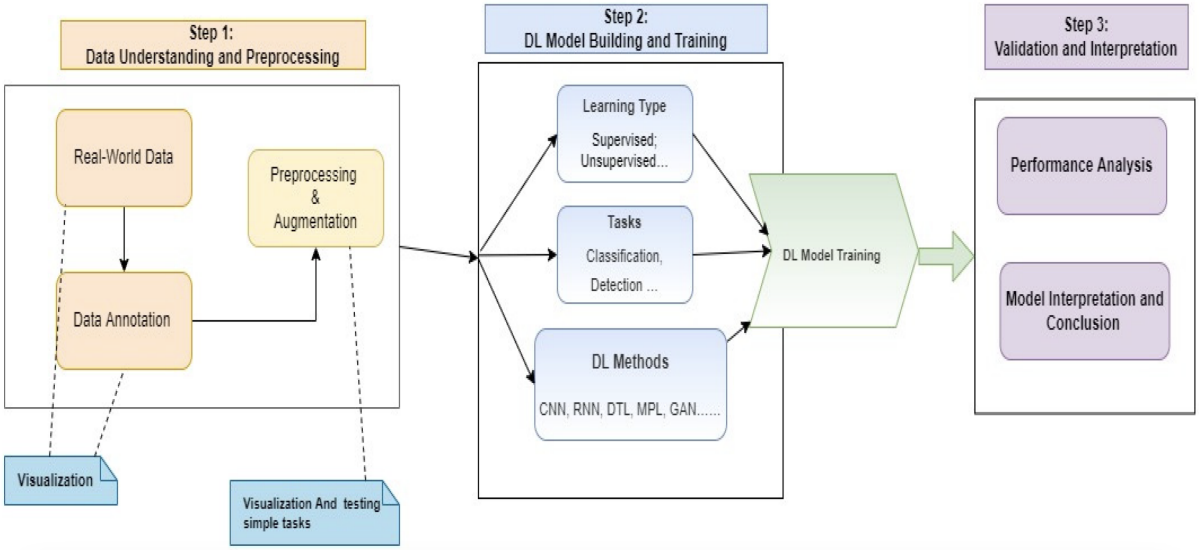


Figure 2.3: A typical DL workflow to solve real-world problems [22].

2.2.1 Artificial Neural Networks (ANN)

Computational models based on the anatomy of intelligent creatures allow ANN to provide a mathematical representation. Its architecture comprises simple processing units storing experimental data.

Designed to replicate the form and operation of biological neural networks seen in the brain, ANN cells known as neurons make up the brain they pass impulses to one another via synapses. Neurons transmit electrical signals to other neurons in response to the messages they receive from their counterparts. Artificial neurons mimic the actions of their biological counterparts by calculating a weighted sum of inputs and subsequently activating their output based on this sum.

If this is greater than a certain threshold, the neuron transmits its signal to its output, which is then received by other neurons.

Thus, weighted synaptic connections link every neuron in the network to other neurons, these weight values either amplify or attenuate the signals moving across the network. Training is adjusting these weight values such that the final response of the network is the right one.

Neural networks usually consist of several layers of neurons. The levels typically consist of

input, hidden, and output neurons. The intricate configuration of neurons in these levels forms a web of interconnected nodes 2.4. As the network learns, its weights are changed to get the result that is wanted by making a certain loss function is minimized.

Activation functions are crucial components of neural networks since they determine whether a neuron should be activated based on its input, hence introducing non-linearity. By including non-linear behavior into the model, they enhance the network's ability to comprehend intricate relationships in the data and accurately replicate complex patterns.

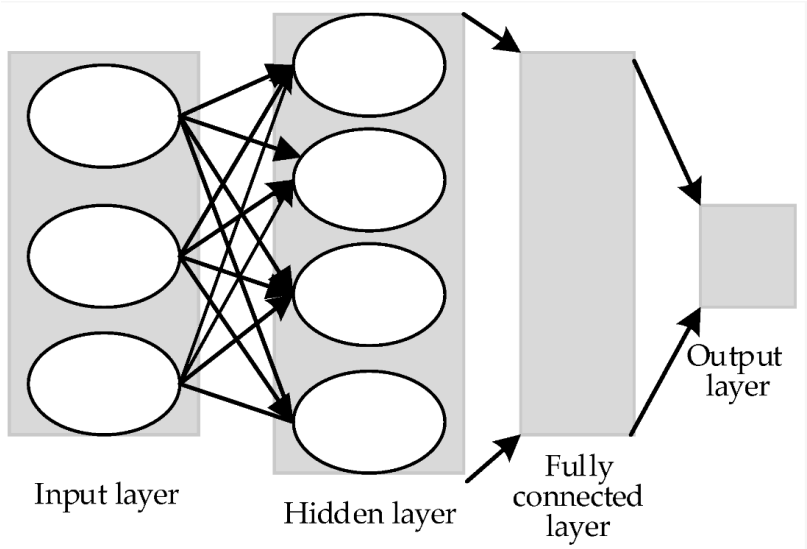


Figure 2.4: ANN structure diagram [10].

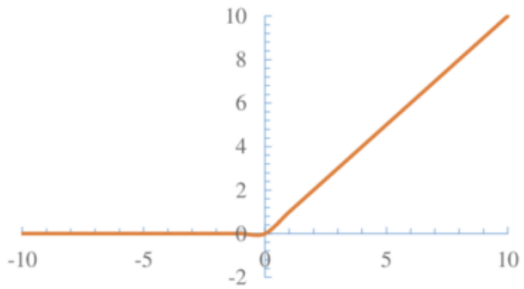


Figure 2.5: ReLu example [10].

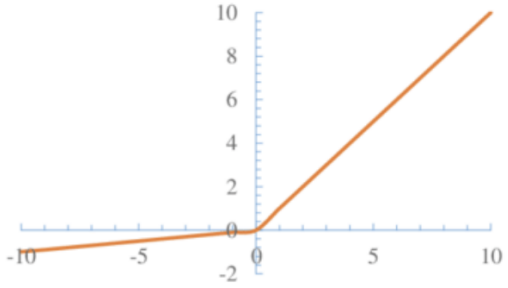


Figure 2.6: Leaky ReLu example [10].

There are several activation functions, some of which are shown in Figure 2.7 and Figure 2.8. ReLu, LeakyReLu, sigmoid, and tanh are the most presently used. The sigmoid function (Equation (2.1)) translates the inputs between 0 and 1 regardless of whether the values are negative or positive:

$$f(x) = \frac{1}{1 + e^{-x}}. \tag{2.1}$$

We often use the activation function ReLU (Equation (2.2)) in the hidden layers. It gives 0 to negative output values and repeats exactly the positive output values from the previous layer.

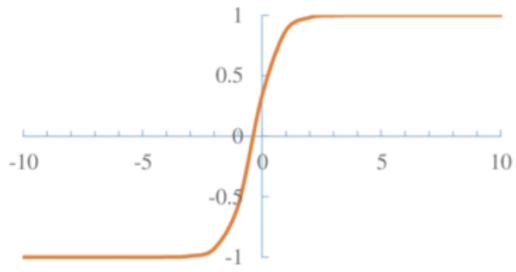


Figure 2.7: Tanh example. [10]

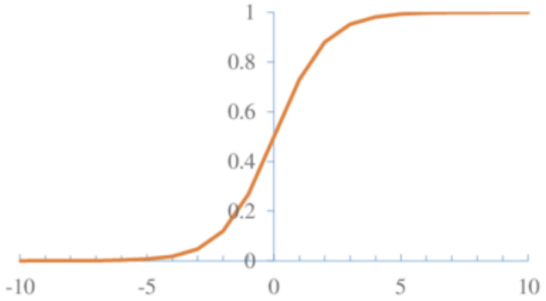


Figure 2.8: Sigmoid example. [10]

$$f(x) = \max(0, x). \tag{2.2}$$

The LeakyReLu (Equation(2.3)) is comparable to the ReLu function, but instead of reducing negative values to zero, it just decreases its extent.

$$f(x) = 1(x < 0)(\alpha x) + 1(x \geq 0)(x). \tag{2.3}$$

The tanh function (Equation (2.4)) maps values between -1 and 1, resulting in increased gradient values and a higher number of weight updates during the training phase.

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}. \tag{2.4}$$

2.2.2 Training an ANN: Process and Techniques

Loss Functions show where the model went wrong when it tried to guess the output values of training samples. The model's main goal is to lower loss to the lowest possible level, which enables it to make more accurate predictions on new data.

For example, in image reconstruction tasks, where the goal is to reconstruct a high-quality image from a degraded or compressed version, loss functions play a critical role.

Mean Squared Error (MSE) is a frequently used loss function in image reconstruction tasks that calculates the average squared difference between the pixel values of the reconstructed and ground truth images. It is defined as:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2 \tag{2.5}$$

The variables x_i and y_i respectively denote the pixel values of the original image and the reconstructed image. The variable n indicates the total number of pixels. Smaller MSE values imply higher reconstruction quality, since the reconstructed image is more similar to the original.

Another loss function, the Structural Similarity Index Measure (SSIM), compares the similarity of reconstructed and ground truth images using both brightness and contrast. It is characterized by a specific definition:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{2.6}$$

where μ_x and μ_y are the mean pixel values, σ_x and σ_y are the variances, and σ_{xy} is the covariance between the original and reconstructed images. Constants C_1 and C_2 stabilize the division. SSIM values range from -1 to 1, with higher values indicating better image quality and structural similarity.

By utilizing these metrics, we aim to comprehensively evaluate the performance of our reconstruction models, ensuring that they provide both accurate and perceptually realistic images.

The weights are modified during training using optimizers based on gradient descent. The computation of the gradient is performed via backpropagation. This optimization-based approach, extensively used in machine learning and neural networks, employs an optimizer to alter the network's parameters, such as weights, with the goal of minimizing loss and increasing accuracy.

The chosen optimizer carries out this change. A hyper-parameter, the learning rate (α) controls the model's ability to adapt itself in reaction to weight fluctuations. Mathematically, it finds the degree of the step followed opposite the gradient throughout the optimization process. High learning rates in the training process could lead to instability. Alternatively, a poor learning rate leads to increased training expenses.

The term batch in the context of the development of DL models refers to a reduced portion of the training dataset that is utilized during a single iteration of the training process. The batch size is the quantity of training instances that are included in a single batch.

When the batch size parameter is set to a high value, additional memory must be allocated to effectively store and manage a larger number of samples simultaneously. Using a smaller sample size, however, can be advantageous in situations where memory is limited, as it requires less memory to process. However, when the model parameters are frequently modified using smaller data sets, the training times may increase and the gradient approximations may become less precise due to the reduced sample size.

During the training phase, it is not possible for the model to achieve a perfect or ideal fit. Underfitting occurs when a simplistic model fails to accurately capture the underlying patterns in the data, leading to poor performance on both new and training data. This phenomenon arises when inaccurate predictions result from a model's inadequate acquisition of knowledge from the training data. On the other hand, overfitting occurs when a model becomes too complex and learns the noise or random variations in the training data. As a consequence, it performs very well on the training data but fails to generalize well to new data. Underfitting and overfitting may both negatively impact the model's performance on unseen data.

The user also has the choice on the number of epochs to apply. This value shows the frequency with which the whole training set is used to modify the model's weights.

Throughout the training process, the model can not fit perfectly or perhaps optimally. Underfitting is the result of a too basic model failing to capture the underlying trends in the data, therefore compromising performance on both fresh and training data. This occurs when erroneous forecasts follow from a model's ineffective learning from the training data. Conversely, overfitting results from a too complicated model learning the noise or random fluctuations in the training data, thereby producing outstanding performance on the training data but poor generalizing to new data. Underfitting and overfitting both potentially compromise the performance of the model on unavailability data.

The model could keep improving its performance on the training data throughout training, but beyond a certain point further training might cause overfitting wherein the model gets excessively customized to the training data and performs badly on fresh data. Early stopping is a method sometimes used to stop this. Early stopping strikes a compromise between underfitting and overfitting by monitoring the performance of the model on an other validation set and terminating the training process when the performance stops increasing.

Batch normalization consists of normalizing the activations of each input variable per mini-batch, such as the enactments of a node from the previous layer, in order to scale the output of the layer. Normalization refers to rescaling data such that it has a mean of zero and a standard deviation of one.

By analyzing the inputs to each layer, it would get closer to achieving the stable distributions of inputs necessary to eliminate the negative effects of the internal covariable shift.

Normalizing the activations of the preceding layer shows that, if at all, the assumptions of the next layer about the distribution of inputs during the weight update will not differ much. This helps the deep neural network training preparation process to be stabilized and accelerated.

For smaller mini-batches that lack an agent distribution of models from the training dataset, the differences in the normalized inputs between training and inference (using the model after training) might result in detectable execution performance differences. This may be remedied with a modification to the approach known as Batch Renormalization, which makes the estimations of the variable mean and standard deviation more consistent across mini-batches.

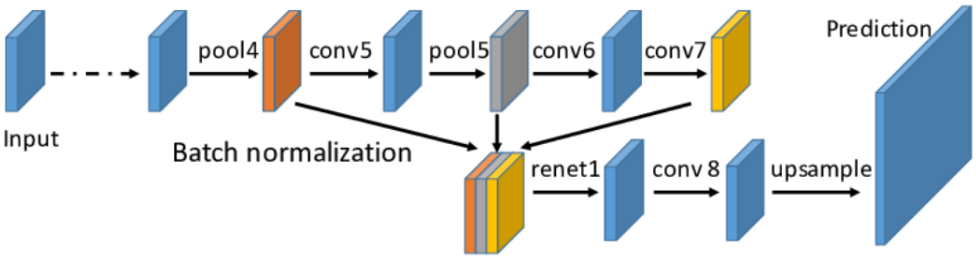


Figure 2.9: Batch Normalization [16].

2.2.3 Convolutional Neural Networks (CNN)

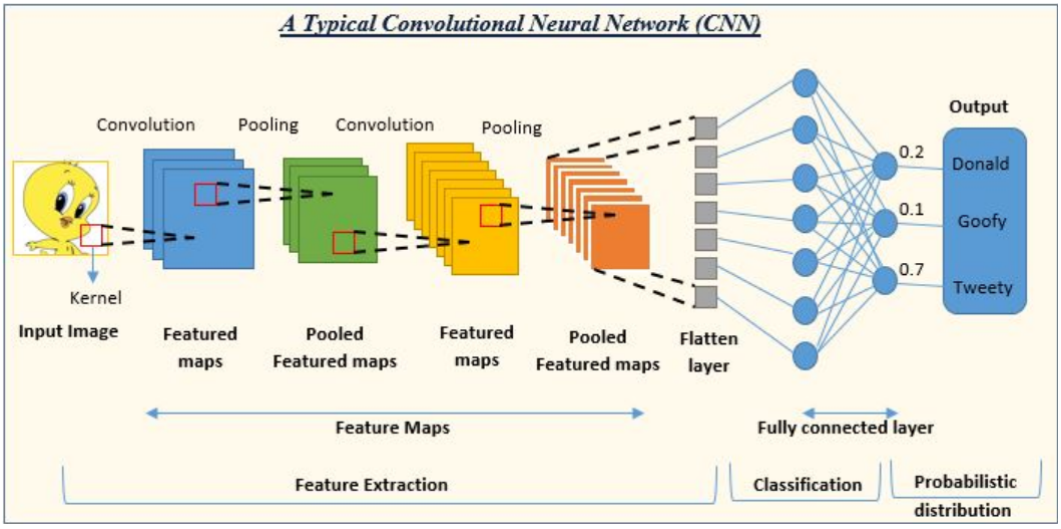


Figure 2.10: CNN [16].

The CNN [6] is a DL technique capable of learning intricate visual patterns. This technique does not need complex preprocessing on the input images due to its capacity to learn their features and map them to extract the most significant ones 2.10.

A CNN consists of several sorts of components. The first layer is the convolutional layer. This sort of layer is used to learn features from the input image, it consists of kernels (filters) paired with strides. These layers are less tightly linked, thus not every input affects every output, and each layer has significantly fewer weights compared to fully connected layers.

One may consider the convolution between the input feature map and the kernel as the sum of the dot product between the translated kernel and the input feature map.

The convolution operation can be mathematically represented as:

$$C(i,j) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} I(i+m, j+n) \cdot K(m,n) \tag{2.7}$$

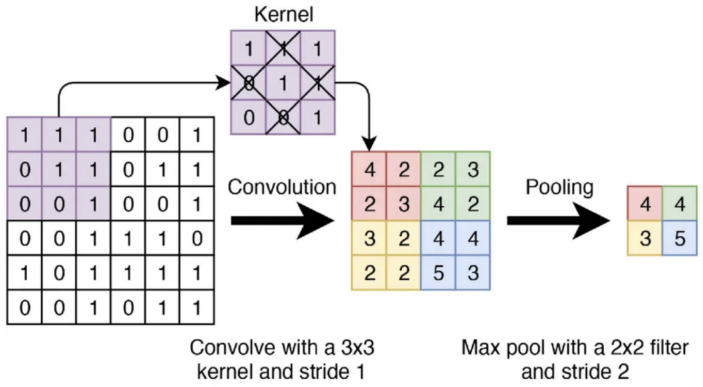


Figure 2.11: Illustration of the convolution operation between the input feature map and the kernel in a CNN convolutional layer [18].

where $C(i, j)$ is the output at position (i, j), I is the input feature map, K is the kernel, and M and N are the dimensions of the kernel.

The strides move the filter over the image, and depending on its size, it may travel one pixel at a time, for instance. In addition, it contains the number of filters in the layer. Additionally, the number of filters in a convolutional layer is equal to the number of output feature maps multiplied by the number of input feature maps.

Conversely, Pooling layers are used to decrease the dimensions of the original data. The pooling layers decrease the width and height of the input volume. This enhances the efficiency of processing and reduces the likelihood of overfitting in the network. Max pooling is often used. Max pooling is used when the highest value inside each patch of the feature map covered by the filter is selected. Average pooling calculates the mean value of each patch, which represents another variant. Pooling layers accelerate training and save memory by reducing the complexity of components and equations inside the network.

When doing upsampling, we use the inverted convolution layer, which reverses the spatial transformation of a conventional convolution layer. To get the dimensions of the output from the convolutional layer, refer to Equation (2.6). For the output size of the inverted convolution layer, consult Equation (2.7), where i stands for the input size, k for the kernel size, s for the number of steps, and p for the padding. Padding represents the number of zero pixels that will be added to the input edge to control the spatial size of the output.

$$\frac{(i + 2p - k)}{s} + 1 \tag{2.8}$$

$$(i - 1) * s + k - 2p \tag{2.9}$$

2.2.4 U-Net

The U-Net architecture, as described in the paper by Ronneberger et al. [20], is specifically designed for semantic segmentation using CNN. It is made up of an encoder and a decoder. The encoder uses convolution blocks and downsampling to extract visual characteristics from the input image.

These characteristics are sent to the decoder, which locates and upsamples them at a higher resolution to enable dense categorization. This decoder is mostly comprised of upsampling and concatenation, with convolution operations occurring on a frequent basis. Figure 2.16 is an example of a simple U-Net design.

The skip connection technique in the U-Net design enhances the flow of information between different levels of the network by establishing direct connections between the encoder and decoder layers. The network improves its capabilities in tasks like image segmentation by preserving low-level information and integrating it with higher-level data. This leads to image segmentation that is highly accurate and reliable. Using skip connections in computer vision applications enhances their effectiveness by facilitating gradient flow during training and minimizing information loss.

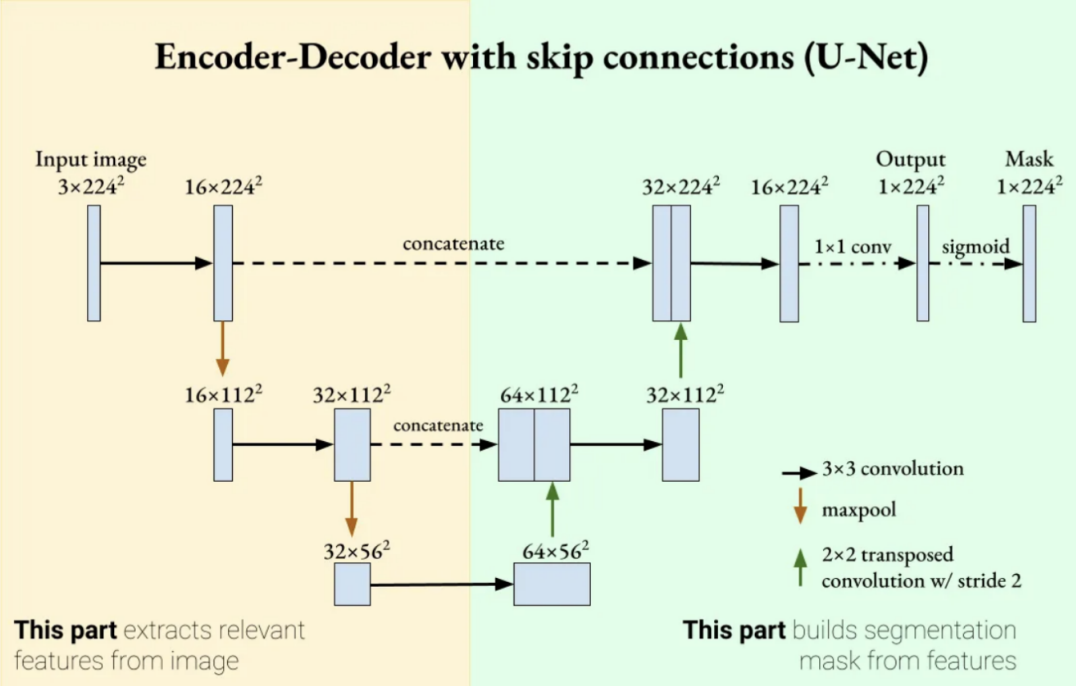


Figure 2.12: U-Net is an encoder-decoder segmentation network with skip connections [20].

The U-Net architecture has significantly transformed image segmentation jobs, especially in scenarios that require segmenting objects inside medical images.

The U-Net design is composed of three essential components: a connection point (bridge), an encoder (downward route), and a decoder (upward path). The encoder, also known as the descending route, consists of convolutional layers that use ReLU activation and pooling operations

to reduce dimensions. Additionally, it might include normalization methods, such as Batch Normalization, to improve the stability of training.

The connection point, or bridge, features an additional convolutional layer that captures contextual information, enabling a smooth transition to the upward path.

The decoder, also known as the upward route, employs transposed convolutional layers to increase the dimensionality. It improves the understanding of the context by combining skip connections and merging the associated feature maps from the encoder. Extra convolutional layers enhance the characteristics, while the last layer utilizes a sigmoid activation algorithm to provide output values ranging from 0 to 1.Improve the steadiness of training.

Activation Function at the Output: The output layer typically employs a sigmoid activation function, especially in binary segmentation tasks, to transform the output into a probability distribution.

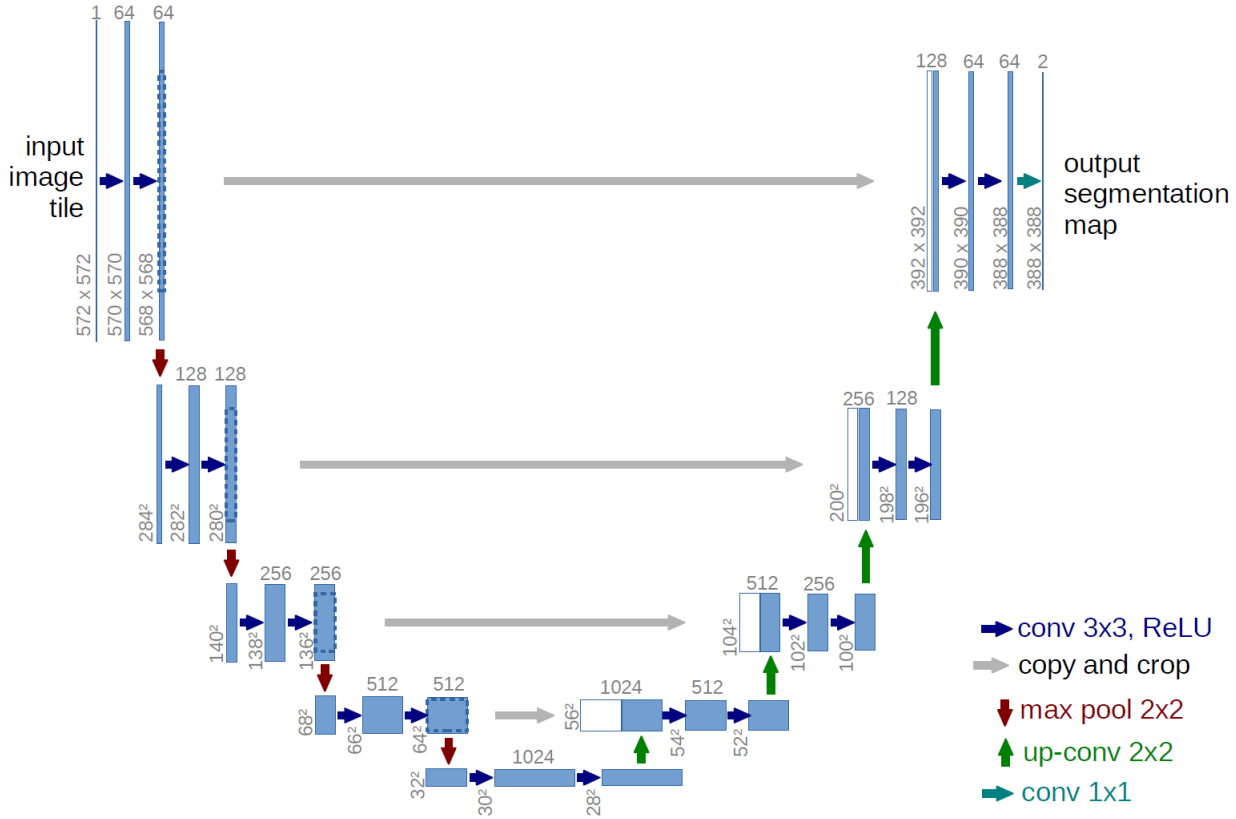


Figure 2.13: U-Net architecture [20].

In the 2D Unet, every layer is composed of two 3x3 convolutions, which are subsequently activated using the rectified linear unit (ReLU) function [8]. The spatial dimensions of the feature maps are reduced by incorporating a 2x2 maximum pooling layer. In the same way, the decoder is comprised of two upconvolution layers measuring 2x2, which are subsequently followed by two convolutions measuring 3x3 and utilizing ReLU activation. To ensure compatibility with the segmentation labels, a 1x1 convolutional layer is employed to modify the number of output

channels.

According to the study by Yi et al. [25], it is recommended to utilize a specific number of channels prior to each max pooling operation in order to minimize data loss. This approach enhances the performance of both the encoder and decoder units by alleviating stress. Repeatedly applying max pooling can result in a significant reduction in the size of the feature map, potentially leading to the loss of crucial information.

Furthermore, batch normalization is implemented prior to each backpropagation iteration to expedite convergence.

For effective training of the 2D Unet, it is essential to have dense volumetric segmentations. However, only labeled two-dimensional segments are used in the training procedure. The model efficiently learns from labeled data due to the use of the weighted softmax loss function, which assigns a weight of zero to unlabeled voxels.

2.3 Image Reconstruction in CT Scanning

Modern CT scanners accurately and fast recreate images from measured data using the FBP technique.

2.3.1 Image Acquisition and Reconstruction Process

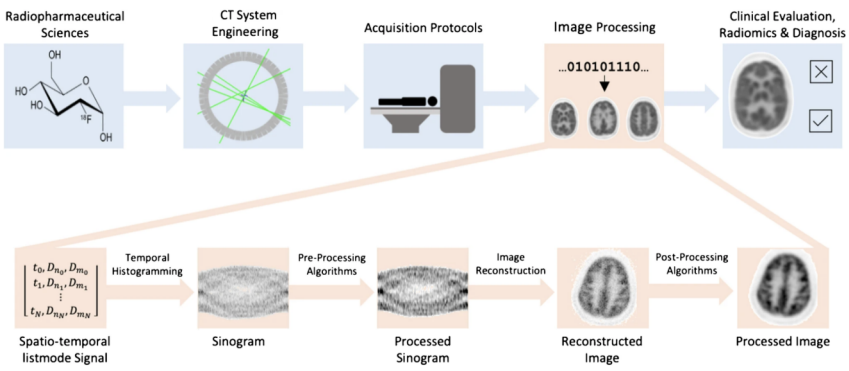


Figure 2.14: DL-based image reconstruction and post-processing methods in positron emission tomography for low-dose imaging and resolution enhancement [3].

Positioning an X-ray source at many points around the body, a CT scan is acquired wherein detectors on the opposite side record the X-rays that pass through. The sinogram produces an image by means of FBP processing. Conversely, if the X-rays are produced from a smaller number of angles, noise and undesired patterns may follow.

Regarding image acquisition, the sinogram shows the unprocessed data obtained during the

CT scan. The X-ray source centers on the person, but the monitors assess the X-ray intensity from many angles as it has gone through the body impression.

A series of images is generated by combining the signals obtained by each X-ray detector. The sinogram is created by presenting these projections in correlation with the location and rotation angle of the detector.

FBP is used to generate an image from the sinogram data. This generates a cross-sectional image that displays intricate anatomical characteristics. The accuracy of the sinogram data directly impacts the ability to reconstruct the image effectively. Reducing the X-ray dosage may alter the signal-to-noise ratio, perhaps leading to imperfections and decreased clarity in the image.

2.3.2 Filtered Back Projection (FBP)

FBP consists of two main steps: data filtering and Back Projection. During the process of filtering, the data is manipulated in order to improve the quality of the image. Back Projection reconstructs the image by projecting data back in the measurement direction, 4.3 for a visual representation of the FBP process. A Ramp filter, a filter is used throughout the process to apply to the projection data to reduce the natural distortion during the back projection stage. The filtered projections are distributed across the original image space to create the reverse projection after a filter addition. This produces the last image that has been rebuilt.

The method starts with the *Radon transform*, which generates a sinogram by converting an image $f(x, y)$ into its projections. The Radon transform mathematically as:

$$R(\theta, t) = \int_{-\infty}^{\infty} f(x, y)\, \delta(x \cos\theta + y \sin\theta - t)\, dx\, dy$$

The projection data (sinogram) as a function of the angle θ and the distance t from the origin $R(\theta, t)$.

Data is altered throughout the filtration process to raise the image quality. In particular, the projection data is filtered typically a Ramp filter to lower the distortion that occurs naturally during the rear projection stage. The filtered projections are distributed across the original image space after application of the filter to rebuild the image. This stage, often referred to as back projection, uses filtered data back along the measurement direction to rebuild the final image.

The FBP process can be mathematically described as:

$$f(x, y) = \int_{0}^{\pi} R(\theta, t) * h(t)\, d\theta$$

where $h(t)$ is the filter applied to the sinogram data, and $*$ denotes convolution.

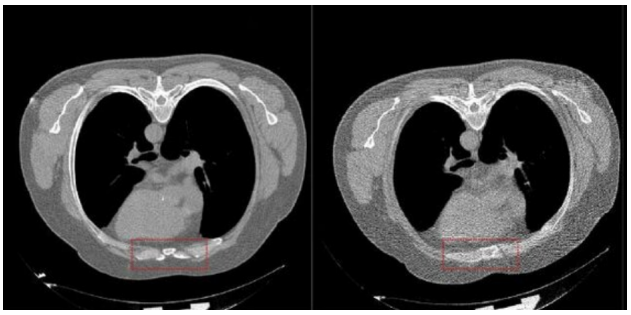


Figure 2.15: FBP. The original image (left) and a CT scan after dose reduction, display increased noise and artifacts(right) [23].

2.3.3 DL for Image Enhancement

Deep neural networks are often used as denoisers for low-dose reconstruction. These networks are taught to identify and eliminate flaws and disruptions that may exist in images reconstructed from low-dose CT scans. These denoising techniques may greatly improve image quality and resolution by using the ability of deep learning models to learn from extensive datasets. Consequently, they mitigate the effects of noise and enhance diagnostic precision.

DL algorithms may improve the parameters used for image reconstruction, therefore producing better and more detailed images especially in demanding imaging conditions. This method is a great tool in developing medical imaging technologies as it helps to reduce the restrictions related with low-dose imaging.

Chapter 3

Literature Review

The most recent developments in using DL techniques with an aim of noise removal and low-dose CT image reconstruction are investigated in this chapter. Low-dose CT image reconstruction is the topic of research aimed on minimizing patient radiation exposure while preserving high-quality diagnostic images. Wavelet-based algorithms, generative adversarial networks (GANs, and CNN among other denoising techniques have been proposed to enhance low-dose CT images. DL methods used for denoising in low-dose CT normally assume that a preliminary backprojection step is performed on the sinogram to get the noisy reconstruction. This first step is crucial as it supplies the input data that the denoising algorithms use to enhance the diagnostic quality and clarity of the rebuilt images.

Gurrola et al. [7] proposed a DL system especially meant for image denoising. This work employs a modified U-Net architecture with residual dense blocks to increase denoising efficiency. Residual dense blocks improve the capacity of the network to store and recycle properties across multiple layers, hence improving the model's power to understand and remove complex patterns.

Residual connections help the network solve the main challenge in deep networks: vanishing gradients by means of integration. The residual dense block is expressed in terms of this equation:

$$\mathbf{F}_i = \mathcal{H}_i(\mathbf{F}_{i-1}) + \mathbf{F}_{i-1}$$

where $\mathbf{F}_i$ is the output feature map of the i-th residual dense block, and $\mathcal{H}_i$ represents the dense block's operations (convolution, activation).

The network proposed in Gurrola et al. [7] aims to efficiently remove noise from images while keeping complex features and structures. The denoising process consists of learning a transformation that turns noisy images $\mathbf{y}$ into clean images $\mathbf{x}$. Training the network helps to reduce the loss function between the expected clean image $\hat{\mathbf{x}}$ and the actual clean image $\mathbf{x}$. This work proposes a combination of perceptual loss and MSE as the loss function:

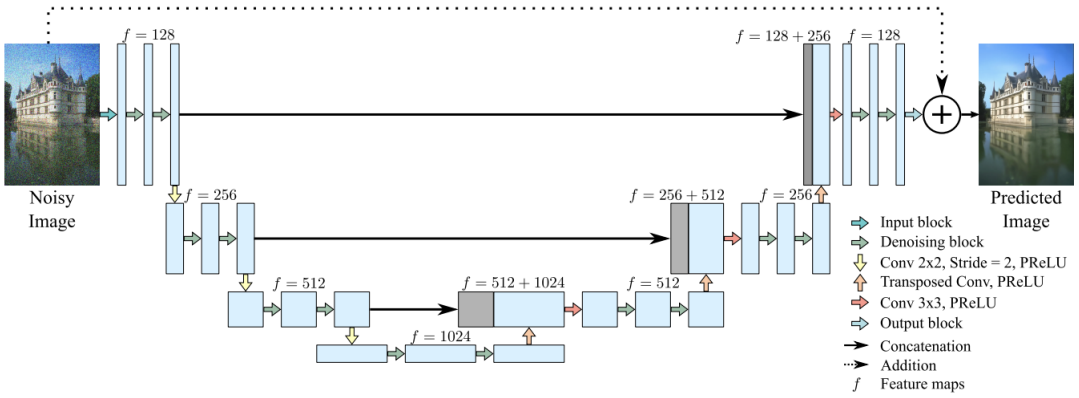


Figure 3.1: Architecture of the network proposed in [7].

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N} \|\mathbf{x}_i - \hat{\mathbf{x}}_i\|^2 + \lambda \mathcal{L}_{\text{perceptual}}$$

where $\mathcal{L}_{\text{perceptual}}$ is a loss function that captures high-level feature differences, and λ is a balancing parameter. The study by Gurrola et al. [7], analyzes image differences in high-level details using the percentage loss metric, to be able to calculate this loss, you first need to collect the details from the pre-trained VGG-19 network, this is a network that was trained on the ImageNet dataset. The network aims to evaluate the differences between denoised images and clean images to minimize noise and at the same time preserve perceptual characteristics such as textures and structures.

Gurrola et al. [7] use several data sets to carry out a detailed evaluation of the proposed model, and managed to obtain very good values compared to Residual Dense U-Net, as the results were better than traditional noise removal algorithms and the metrics used to evaluate image quality as Peak to Signal Noise (PSNR) and SSIM.

The study Liu et al. [13] investigates the removal of the noise from low-dose CT images using CNN. The objective of the study is to use techniques from the U-Net architecture to help remove noise from images. These techniques often have difficulty removing noise from the image and at the same time maintaining its quality. The U-Net CNN was modified in the study by Liu et al. [13] to be able to resolve these issues.

To be able to choose specific regions of the images, the proposed U-Net network uses attention methods to differentiate between noise and crucial elements, that is, the network uses both channel attention and spatial attention processes. Channel-based attention gives the network more priority to the channels with better information, while spatial attention during processing gives the crucial parts of the image greater emphasis [13].

MAGUNet's architecture, seen in Figure 3.2, is organized around many significant components. The network includes a multi-scale feature extraction block that captures low-level visual characteristics. An AGB selectively emphasizes significant image areas, hence improving feature

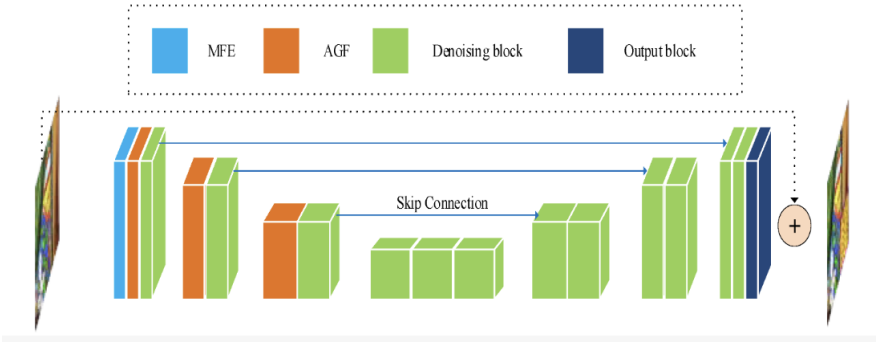


Figure 3.2: Architecture overview of MAGUNet, comprising a multi-scale feature extraction block (MFE), attention-guided filter block (AGB), and residual denoising block (RDB) [13].

representation. Resilient learning is combined in the RDB to efficiently lower noise while maintaining image features. By use of both the encoder and decoder sections, these elements are linked and enable MAGUNet to have complete feature extraction and noise reduction capacity.

Furthermore included into the network is a residual learning technique, which substitutes residual noise instead of the clear image directly. This approach improves the ability of the network to reconstruct complex aspects in the denoised images and simplifies learning. Furthermore included in the network are multi-scale feature extraction blocks to gather features at different resolutions, therefore enhancing its performance in noise reduction and image detail preservation.

The experimental results suggest that the attention-based U-Net outperforms a variety of state-of-the-art denoising methods, particularly in the areas of noise reduction and the preservation of intricate features. The denoised images produced by this network exhibit enhanced resolution and structural similarity in comparison to previous methodologies.

Aiming to improve the CNN technique for CT reconstruction from low-dose data, Heinrich et al. [9] suggested two multi-layer CNN architectures for denoising CT images. Comprising three blocks of five-by-five convolutional filters, the ResFCN is a fully convolutional network. Whereas both models use an input image residual connection to aid learning, the ResUNet is trained using 10 convolutional blocks organized in a multi-scale form. Using the XCAT Phantom [9] for realistic simulations to generate training images, Heinrich et al [9] showed that the ResUNet approach yields the most promising outcomes, achieving a peak signal-to Noise ratio of 44.00 dB, compared to 41.79 dB for the ResFCN [9].

Figure 3.4 shows the Residual U-Net model, where we can visualize three types of different colors that mean three types of different resolution scales (green for high resolution, yellow for medium resolution, and red for low resolution) with the aim of achieving a more efficient way of observing the characteristics of medical images. This method can maintain structural information in CT images while allowing complete noise reduction, which is necessary for an accurate diagnosis. Furthermore, radiologists evaluated the denoised CT images produced by these structures, where there was an improvement in image sharpness, noise reduction, and better visibility of the structures using as metrics the PSNR, and subjective assessments.

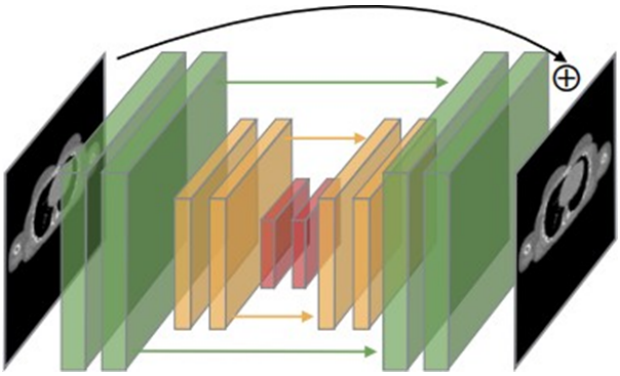


Figure 3.3: Architecture overview of Residual U-Net [9]. The network leverages multi-scale feature extraction to capture detailed image priors across different resolution levels.

Using a generative model grounded on a GANs technique, Zhang et al. [26] seeks to lower noise in low-dose CT. Unlike other models, this one uses a DU-GAN architecture wherein the two discriminators U-Net based operate in the image and gradient domains while the generator is in charge of noise reduction. Whereas the other guarantees that the produced CT images in the image domain closely match the source images, one discriminator improves clarity at the borders and lowers banding artifacts in the gradient domain. This dual-method technique effectively lowers noise in the reconstructed images while preserving sharpness, hence improving the denoised image quality generally.

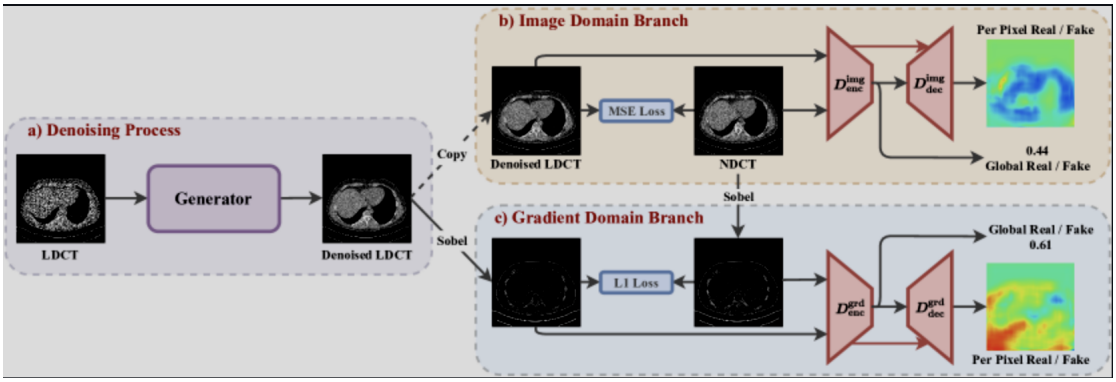


Figure 3.4: Overall framework of the proposed DU-GAN[26].

The dual-domain approach within the DU-GAN model allows one to preserve important structural components in denoised images, which could be destroyed by standard post-processing methods. Using U-Net-based discriminators that combine pixel-level properties with general structure, the model efficiently lowers noise. This guarantees that CT scans are accurate for diagnosis even at normal doses, therefore conserving important information.

The study by Gu et al. [5] talks about quantum noise in low-dose CT images and how to combat this noise, as it is negative for image accuracy. Gu et al. [5] talks about an approach to combat this problem that aims to improve image quality and reduce noise more effectively.

The results of the evaluation revealed that the produced images by the suggested method

had considerably higher quality than low-dose CT images and baseline CycleGANs. This work employed CycleGANs, a class of GANs that enables image-to-image translation without depending on paired data. Noise was essentially disentangled from extremely low-dose CT images by including a wavelet transform into the CycleGANs architecture. The wavelet transform aids in the breakdown of the image into various frequency components, therefore enabling the CycleGANs to concentrate more specifically on denoising certain regions without sacrificing significant structural information. This approach may be a very important denoising platform for low-dose CT.

The proposed method utilizes a adversarially trained network in combination with a sharpness-detecting network. More precisely, the network is engineered to acquire knowledge and implement a process that distinguishes random interference from the valuable information in the CT scans. The adversarial network is trained within the framework of a GANs, where the primary goal of the generator is to generate denoised images that are indistinguishable from clean images. Conversely, the objective of the discriminator is to differentiate between denoised images and clean ones. Furthermore, a network is trained to discern sharpness in order to preserve the intricate details of edges by comparing the sharpness of the generated images with that of the reference images.

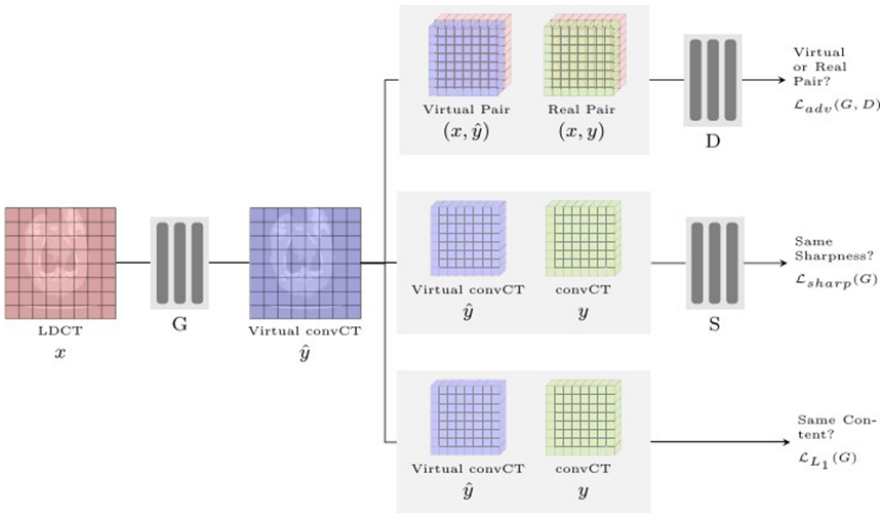


Figure 3.5: Overview of SAGAN [5].

Gu et al. [5] propose the Self-Attention Generative Adversarial Network (SAGAN) as a means to improve image quality. Conventional denoising methods that use non-linear filtering often encounter the problem of edge blurring. This occurs because these methods tend to average nearby pixels or comparable patches, resulting in a reduction in the sharpness of edges [25]. Under circumstances with high levels of noise, the noise itself has the ability to create false edge like formations. Neural network-based approaches may encounter similar difficulties when optimizing the differences between produced images and ground truth on a pixel-by-pixel basis, since they may seek to blend features together in order to reduce overall disparities.

SAGAN addresses the challenges of incorporating self-attention processes, allowing the model

to focus on long range connections and the full context within the image. SAGAN has specific features that provide uniformity in different areas of the image, which will result in the retention of intricate elements and greater image clarity.

The adversarial loss that is caused by theGANs discriminator results in a more unique image, but at the same time, the adversarial loss does not guarantee the sharpness of the reconstructed images, especially in areas with low contrast. To solve this problem, the study by Gu et al. [5] created a separate network specifically to detect sharpness and as part of SAGAN training, the generator creates a virtual convCT which is then passed to the sharpness detection network and finally does a comparison of the sharpness map with the map of the reference image.

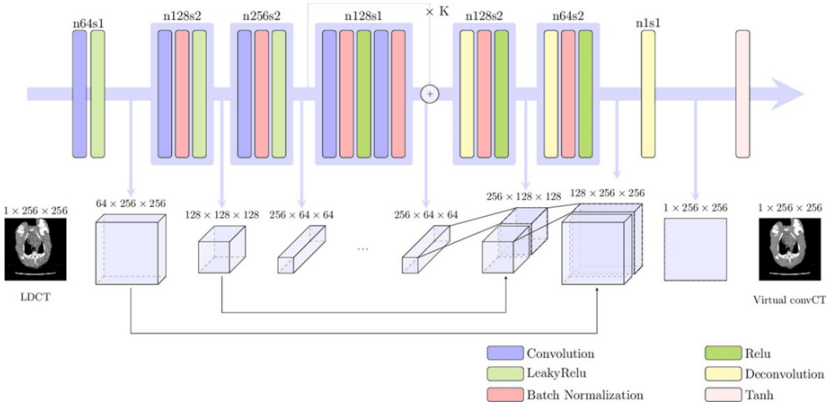


Figure 3.6: Proposed generator of the SAGAN. The residual block in the center of the network is repeated K times, and K was chosen as 9 for the experiment [25].

The final feature map of the sharpness detection networks, the sharpness map, is a gray-scale image where pixel intensity indicates the local sharpness of a specified region. The mean square error is computed between the two sharpness maps, and this error is fed back through the generator to update its weights.

Details may be created in SAGAN utilizing indicators from all feature locations. In addition, the discriminator may verify the consistency of extremely detailed features in distant areas of the image. Recent research has shown that generator conditioning impactsGANs performance. Using this understanding, it was possible to apply spectral normalization to theGANs generator and see an improvement in training dynamics.

The self-attention module complements convolutions and facilitates the modeling of long-range, multi-level relationships between visual areas. The generator's self-awareness enables it to produce images in which fine details at every place are meticulously matched with fine features in the image's remote regions.

In addition, the discriminator may apply intricate geometric restrictions on the global image structure with greater precision. In addition to self-awareness, current findings about the relationship between network conditioning andGANs effectiveness were also included.

Gu et al. [5] suggest that to produce consistent objects/scenarios, the proposed SAGAN

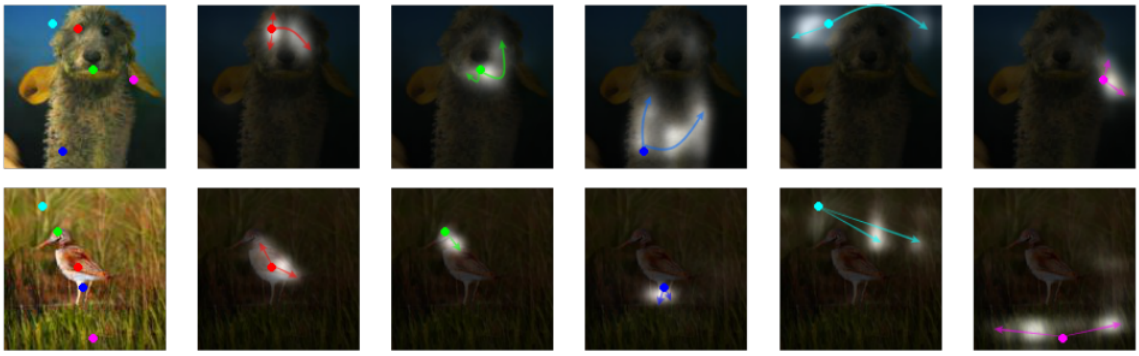


Figure 3.7: SAGAN [5].

generates images by exploiting complementary information in remote areas of the image rather than local regions of fixed shape. The first image in each row depicts five sample query sites with color coded dots. The last five images are attention maps for these query locations, with color coded arrows highlighting the most visited areas.

Wavelet-based noise reduction techniques are often used to remove noise from CT images. It is possible to describe the techniques using the following steps: the wavelet transform breaks apart the noise-contaminated CT image into wavelet coefficients that at different frequencies capture visual information. High-frequency noise components are found and deleted using a thresholding mechanism. By inverse wavelet transform on the thresholded coefficients, the image is recreated. This procedure produces a denoised image with most of the original information preserved and noise level lowered.

Noise reduction is highly influenced by the choice of the wavelet function and thresholding so, the wavelet function is chosen dependent on the image properties and noise. Noise from the image is eliminated by thresholding function modification.

Wavelet-assisted noise reduction is also a technique that is used to improve the quality of CT scans, that is, it preserves the characteristics of the images while removing noise from them.

Our study of the literature primarily focuses on investigating the techniques of DL used to remove noise from CT images. It is essential to understand the link between modern DL approaches and classic wavelet-based methods. Wavelet-based methods have shown promise for both noise reduction and retention of important visual characteristics. Still, these methods mostly rely on human choice of wavelet functions and thresholding methods. Still, powerful machine learning techniques like CNN and autoencoders help to streamline the identification of traits and acquisition of the most effective representations straight from the input. Recent studies by Smith et al. [21] and Lee et al. [12] show that gathering complex patterns and noise characteristics from large datasets DL models has the potential to exceed conventional approaches. Sometimes more than wavelet-based techniques, these models can efficiently lower noise in images with different degrees and patterns of noise.

We might understand the extent to which DL methods surpass and improve wavelet-based

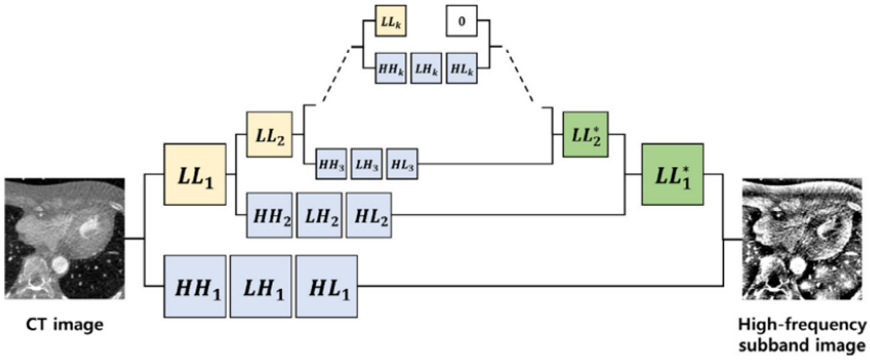


Figure 3.8: The process for wavelet-assisted noise disentanglement by extracting high-frequency sub-band images from a CT image using k-level wavelet transform and inverse wavelet transform [5].

denoising capacity by comparing these two methods. From manually created solutions to data-driven, automated systems using DL capabilities, this comparison also highlights the developing aspects of image processing techniques in medical imaging.

Using a deep CNN, Kang et al. [11] suggest a technique to apply to the wavelet transform coefficients of low-dose CT images.

Using a directional wavelet transform to capture the directional component of artifacts and exploiting intra and inter band correlations, their deep network can reduce CT specific noise efficiently. Moreover, the CNN is developed with a residual learning architecture for quicker network training and improved performance. Experimental findings demonstrate that the proposed technique efficiently eliminates complicated noise patterns from CT images produced from a decreased X-ray dosage.

Additionally, Kang et al. [11] demonstrate that the wavelet-domain CNN is superior to current methods for removing noise from low-dose CT scans.

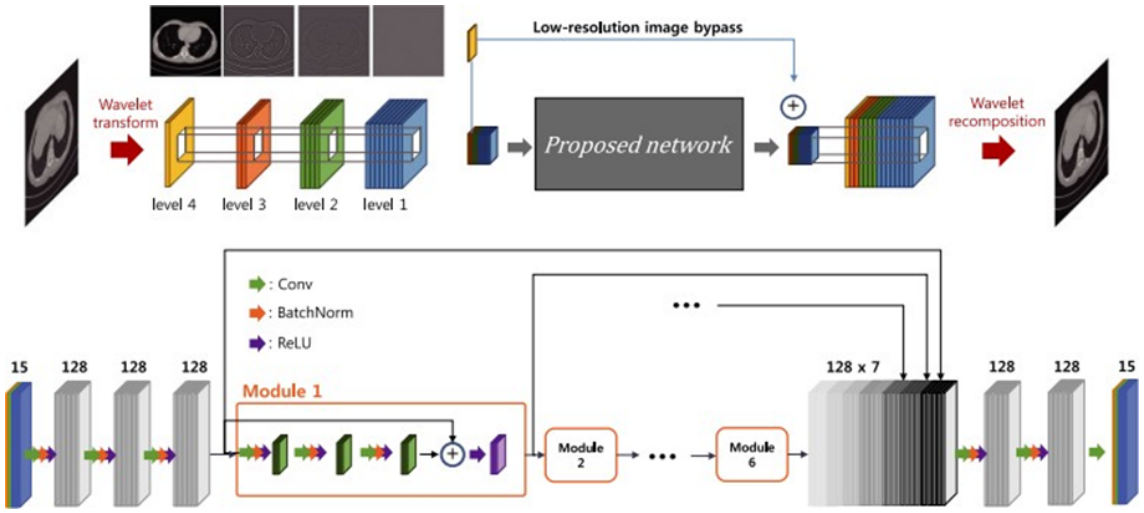


Figure 3.9: Proposed deep CNN architecture for wavelet domain denoising [11].

3.1 Discussion

An overview at the most recent progress and problems in making contrast CT scans and lowering the amount of noise in low-dose CT images shows big steps forward and problems that still need to be solved. Convolutional designs have shown big gains in image clarity and reducing noise, especially in deep-learning models. The use of dual-domain techniques, latent learning, and attention processes makes these models much better. Wavelet-based denoising methods are another option that works very well for dealing with complex noise patterns.

Various patterns can be identified in the literature that have been analyzed as trends in the field.

Liu et al. [13] illustrate in their study the present trend toward increasing the model's concentration on important places within images is the integration of attention processes, thereby enhancing both detail preservation and denoising. Residual learning similarly solves the vanishing gradient problem a challenge tackled in many studies so maintaining fine features in denoised images.

Known as the Dual-Domain Approach, Zhang et al. [26] have produced a significant study emphasizing the need of analyzing both the image and gradient domains to efficiently remove noise while maintaining edge features. This method shows how combining many approaches, including CNN) with wavelet transformations, may provide hybrid methods that use the strengths of each to get outstanding results.

However, due to their great computational complexity, the significant processing resources needed by DL models especially those including attention processes cause problems for real-time applications. Moreover, the need of consistent solutions applicable in a wide range of clinical situations is underlined by the fact that the applicability of denoising models to different datasets and noise levels is still a subject of discussion.

Wavelet-based denoising approaches need meticulous selection of the wavelet function and thresholding methodology, which must be suitably adjusted for best outcomes. One of the primary difficulties in this field is striking the optimal balance between reducing noise and preserving details. Excessively low noise levels might potentially result in the loss of vital diagnostic information.

Ultimately, the progress in DL and wavelet-based denoising methods shows potential for enhancing the quality of low-dose CT images. Nevertheless, further investigation is required to tackle the computational difficulties, improve the capacity to apply findings to a wider range of situations, and fine-tune the trade-off between reducing unwanted interference and maintaining accurate information. The advancements in DL and wavelet-based denoising techniques have the potential to improve the quality of low-dose CT images. However, further research is necessary to address the computational challenges, enhance the ability to generalize results to other scenarios, and optimize the balance between minimizing unwanted interference and preserving

correct information. Although the use of new tactics and the integration of various approaches are anticipated to drive progress in this field, there is still a significant absence of ways that successfully integrate existing side information to enhance reconstruction.

In Table 3.1 we can see a small summary of the performance of each algorithm considered in this literature review.

Table 3.1: Summary of Papers on Image Denoising Techniques

Paper	Objective	Approach	Results	Data Used
Gurrola et al. [7]	Image denoising.	U-Net with residual dense block. Loss function combining MSE and perceptual loss.	The residual dense U-Net works better than both simple and complex neural network-based denoising methods.	Natural images dataset
Liu et al. [13]	Low-dose CT denoising.	U-Net with channel and spatial attention mechanisms. MFE, AGB, and RDB blocks.	Enhanced noise reduction and enhanced preservation of details.	Low-dose CT images: 100 patients, 10,000 CT scan slices
Heinrich et al. [9]	CT reconstruction.	ResFCN and ResUNet. Multi-scale convolutional blocks.	PSNR = 44.00 dB (ResUNet), 41.79 dB (ResFCN).	Low-dose CT data: 50 patients, 5,000 CT scan slices
Zhang et al. [26]	Low-dose CT denoising.	Utilize DU-GAN with U-Net-based discriminators. Utilizing a dual-domain technique.	Enhanced noise reduction while maintaining the integrity of features and structure.	Low-dose CT images: 120 patients, 12,000 CT scan slices
Gu et al. [5]	Quantum noise reduction in CT.	SAGAN with self-attention. Sharpness-detecting network.	Enhanced image quality and sharpness.	Low-dose CT images: 80 patients, 8,000 CT scan slices
Kang et al. [11]	CT noise reduction.	Residual learning deep CNN with wavelet transform.	Superior noise pattern removal.	Low-dose CT data: 60 patients, 6,000 CT scan slices

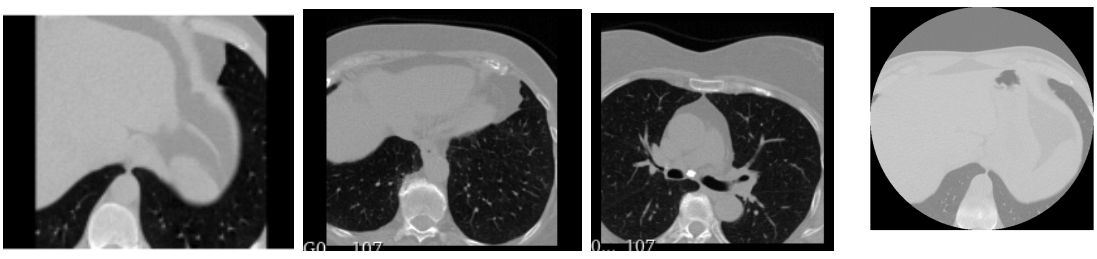
Chapter 4

Methods

The objective of this chapter is to describe the materials used in this thesis and to describe in detail the methods proposed to reconstruct contrast-enhanced CT scans from low-dose measurements. In particular, this chapter will describe solutions based on the U-Net architecture and propose a model to reconstruct CT images that does not use side information, along with two different ways to incorporate side information coming from non-contrast CT.

4.1 Materials

The current research utilizes datasets derived from the Orca Score platform inside the Grand Challenge [24]. Training and testing sets have already been established out of the data. The training set consists of thirty-two patients overall, and the testing set consists of thirty-two patients also. Every patient is split further into non-contrast and contrast CT scans. Figure 4.1 shows for every train and test set four different forms of CT (several machines).



Each patient's contrast and non-contrast information is stored in pairs of `mhd` and `raw` formats that include all the CT data. The `mhd` file functions has a header file that holds all of the image's information, while the `zraw` file contains the compressed image data. This dataset's slice thickness is separated into four groups, with eight patients in each group.

The contrast images of the first group of patients had a slice thickness of 0.625mm, while the non-contrast images had a slice thickness of 2.5mm. The slice thickness of the other three groups

in non-contrast images was consistently 3mm. However, the thickness varied in contrast images. The slices in the second group had a thickness of 0.250mm. Conversely, the third group had a thickness of 0.45 millimeters, whilst the fourth group had a slice thickness of 0.40 millimeters.

By using Slicer [2], we converted all of these files into slices in the `dicom` format. The `dicom` file provides an image of a patient, along with pertinent patient information. This file further includes data on the patient's whereabouts and other metadata. Table 4.1 shows the ultimate quantity of slices obtained from the Slicer.

Folder	**Nº Patients**	**Type of CT**	**Nº Slices**
Train	32	Contrast	10685
Train	32	Non-Contrast	1540
Test	32	Contrast	10614
Test	32	Non-Contrast	1551

Table 4.1: Orca Data.

4.2 Data pre-processing

Prior to using these images in the three models being evaluated for this dissertation, it is essential to do pre-processing to enhance the overall quality of the models. The first pre-processing stage included aligning contrast and non-contrast CT images. This was essential because to the disparity in the quantity of images between the two kinds of scans. The process of filtering the images guaranteed that there was a one-to-one correspondence between contrast CT scans and non-contrast CT scans.

In this thesis, two alignment pre-processing approaches were examined: manual selection, which aimed to choose the closest contrast slice for each non-contrast slice, and image registration, which included pairing the slices together.

The first choice was decided based on the Orca Data, since it served as the primary dataset for training the models. Several non-contrast slices were excluded because there was no adjacent contrast slice available. According to the data shown in Table 4.1, the number of slices is now the same for both contrast and non-contrast CT scans, but the total number of slices is smaller.

Folder	**Nº Patients**	**Type of CT**	**Nº Slices**
Train	32	Contrast	1406
Train	32	Non-Contrast	1406
Test	32	Contrast	1422
Test	32	Non-Contrast	1422

Table 4.2: Summary of the Orca Dataset's Paired Filtration.

The datasets include images with a resolution of 512 by 512 pixels, which means that analyzing

the pixels will require more memory and take longer to train. If the resolution were lower, it would be possible to save more memory, but it would have the disadvantage of losing some details. This reduction of the image resolution by half, means that images with a resolution of 256 by 256 pixels were used. It was completed when the resizing function provided by the OpenCV package was employed, as it allowed choosing the desired interpolation function, and this interpolation was identified as INTER AREA.

Given that pixel values for CT scans are initially expressed in Hounsfield Units (HU), which may exceed the range [0, 255], we standardized the images to ensure uniformity. This normalization guarantees compatibility with the Orca dataset.

In the Orca dataset, after obtaining the images with the same range of pixel values, it was necessary to transform these pixel values to a smaller range so that they could be used in the model, as a larger range would mean in greater waste of resources and computing. So it was decided to set the pixel range to [-1, 1]. Finally, a function was developed that is used in linear conversion to transfer the data to other ranges.

To avoid needing to use the data sets whenever we carried out this pre-processing, that is, for ease of use, both sets were saved separately in the NumPy (.npz) format. The NumPy format is a method used to store NumPy arrays, and as the images have this format, it was only necessary to generate a single array that contains all the images. In the case of the Orca dataset, two arrays were used, one for images without contrast and the other for images with contrast. Finally, to be able to store the arrays, the `savez compressed` function of NumPy was used, which has the functionality of saving both arrays in the same file and then loading the data set, it is enough to re-separate the arrays to obtain images with and without contrast.

4.3 Architectures and Training

This section will provide a detailed explanation of the U-Net architecture, with a specific emphasis on the three primary models examined in this thesis:

- **Reconstruction without side information:** The objective of this model is to generate CT scans using only the input data, without relying on any further information from non-contrast images.
- **Reconstruction with side information using 2 channels:** This strategy utilizes two channels to input side information into the model. The contrast-enhanced CT scans are processed by one channel, while the non-contrast CT images are processed by another channel simultaneously. This allows the model to utilize information from both sources in order to enhance the accuracy of reconstruction.
- **Reconstruction with side information using 2 encoders:** One focused on handling the contrast-enhanced CT images and another on the non-contrast CT images. Combining

the encoded characteristics of both encoders and transmitting them via a common decoder produces the final CT image. This approach aims to gather and combine more complex information from both image databases.

4.3.1 Computation of CT Measurements and Filtered Backprojection

Before describing the three models that will be used in the thesis for image reconstruction, it is first necessary to describe how the CT measurements were generated, such as the sinogram. The sinogram was created by the Radon transform, as it generates the projection data necessary for the reconstruction of the CT image, being calculated using the `radon` function included in the `scikit-image` library. The background chapter contains a thorough explanation of the sinogram along with the Radon transform formula and illustrative figures.

Thus, it was possible to determine the lines that were preserved in the Radon transform to ultimately replicate low-dose data. To use smaller projection angles, the radiation dose was reduced, and the low-dose data were simulated by retaining only a subset of the projection angles, distributed equally over 180 degrees. Different numbers of angles (64, 128, 256) were considered for reconstruction by FBP in order to evaluate the quality of image reconstruction and we can conclude that the greater the number of angles, the greater the image quality, knowing that the image is made up of 512 angles.

Using the `scikit-image` library, more especially with the `iradon` function, FBP was also carried out. This operates the inverse Radon transform to recreate the image from the sinogram. Using a large number of projection angles specifically, 512 lines filtered backprojection was used to acquire the "ground truth" images for our studies, therefore guaranteeing a high-quality reconstruction that acts as a baseline for assessing the models.

4.3.2 Reconstruction without side information

The first model that will be discussed is a more traditional model, which consists of an encoder-decoder structure with skip connections. To capture high-level images, the encoder neatly reduces the spatial dimensions of the input CT image while at the same time increasing the number of feature channels. The decoder now behaves differently, i.e., it reverses the encoder procedure, using the extracted features to restore the image to its initial dimensions. And as the model name says, it uses only the input CT image, without adding additional data from non-contrast CT images.

The encoder component of the U-Net architecture has 4 convolutional layers and each of these layers is followed by a ReLU (Rectified Linear Unit) activation function, and in each of these layers, the number of feature maps calculated: 64, 128, 256 and 512, respectively. The spatial dimensions of the convolutional kernels are 3×3. Max-pooling layers with size 2×2 are used to perform down-sampling after each convolutional layer, while the decoder component

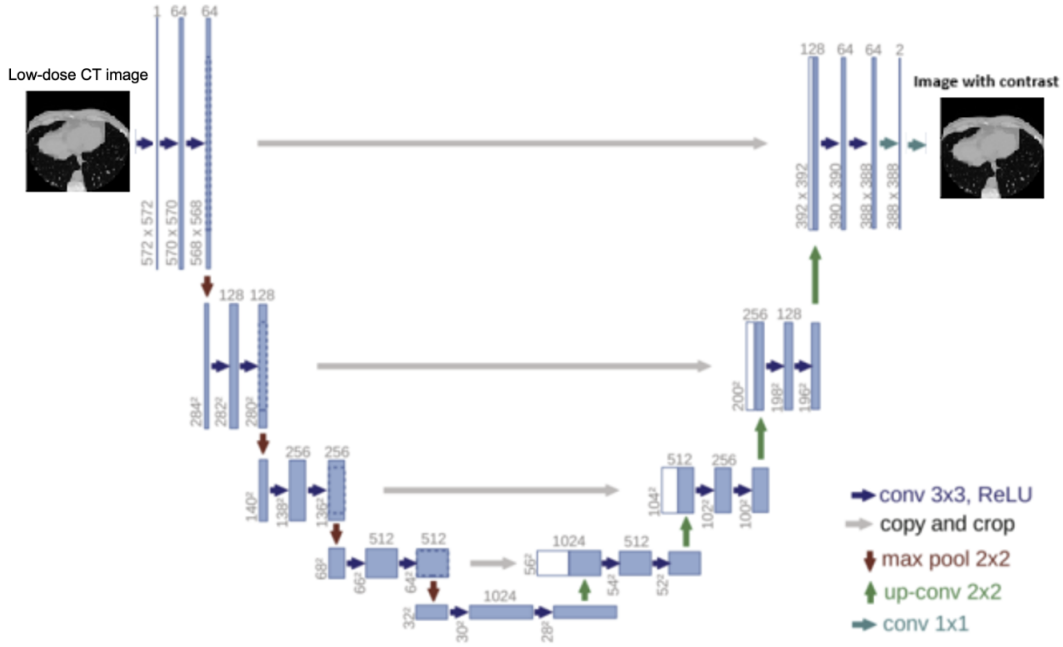


Figure 4.1: schemes of the reconstruction without side information [27].

replicates this ordering using convolutional layers followed by up-sampling layers. To help restore the spatial information that will be lost during the down-sampling process, skip connections are used between the corresponding encoder and decoder levels. Finally, to train the architecture, a loss function is used, such as the MSE, which serves to reduce as much as possible the differences between the reconstructed images and the original CT images.

4.3.3 Reconstruction with side information using 2 channels

In the second model, it is already a more sophisticated model in relation to the U-Net design, as it includes auxiliary information by adding a second input channel. The model has two channels, one of the channels analyzes tomography images with contrast and the other channel analyzes tomography images without contrast, both channels will be placed in the same encoder so that the model can learn from the two images in simultaneously.

The encoder in this architecture, as previously mentioned, is similar to a conventional U-Net network, but it was adapted so that it can have two input channels in the same encoder. The convolutional layer works with the characteristics that are combined from both images, this means that when the network is using data with contrast, it has an auxiliary image without contrast. The decoder no longer changes anything, it remains the same as the conventional U-Net network, using the merged features to recreate the final tomography image.

To conclude, this two-channel technique is used to help the model improve CT images, getting more detailed and accurate images.

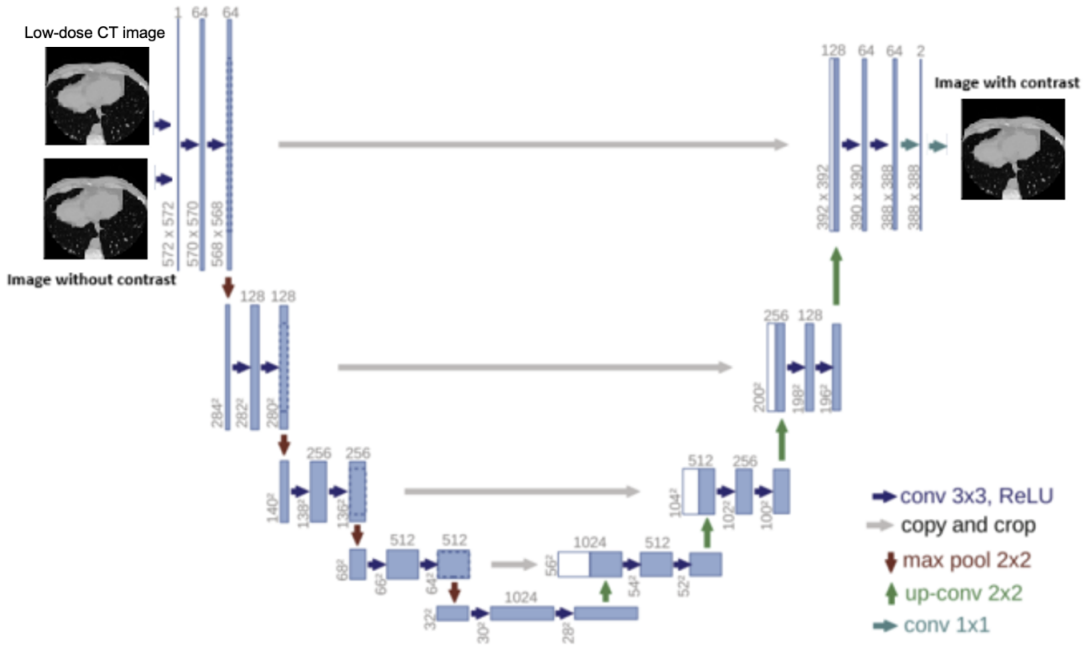


Figure 4.2: schemes of the reconstruction with side information using 2 channels [27].

4.3.4 Reconstruction with side information using 2 encoders

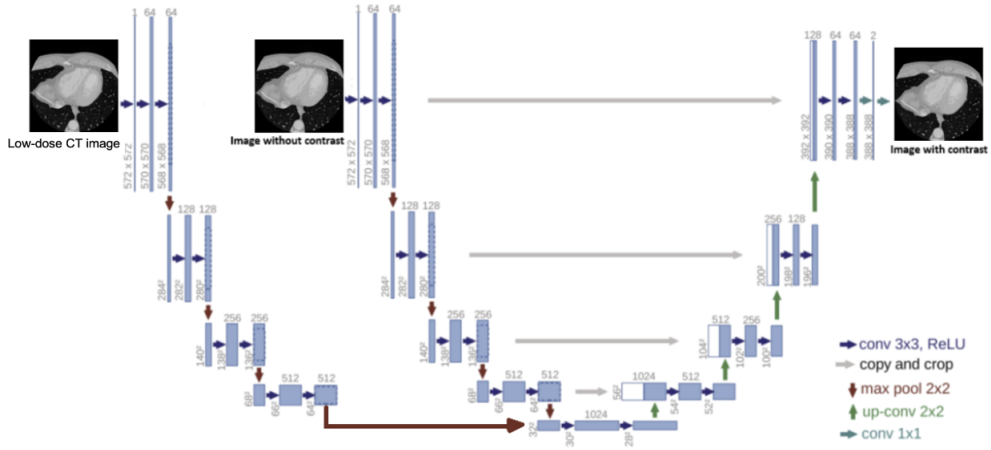


Figure 4.3: schemes of the reconstruction with side information using 2 encoders [27].

Finally, the third model is a more complex model, which has two different encoders, one of which will have images with contrast and the other images without contrast. Each encoder is used to extract distinct features from your image. After that, the features from both encoders are combined and sent to a common decoder, which will ultimately generate a final reconstructed CT image.

Each encoder in the project has convolutional layers with Rectified Linear Unit (ReLU) activations and max-pooling that are similar to the conventional U-Net encoder. However, these encoders are customized to suit their respective input types. Following the encoding process, the feature maps that are generated by both encoders are combined and inserted into the decoder, if followed, the decoder uses the similar architecture to previous models, where it will include up-sampling and convolution layers to reconstruct the image.

Finally, the use of a dual encoder allows the model to more effectively gather more complex data from CT images with and without contrast, which has improved performance in the reconstruction process.

You can also remark the fact that one of the main reasons for the success of the dual encoder solution is the fact that contrast-enhanced and non-contrast CT scans are not exactly aligned. In this sense, combining the information coming from the two of them directly from the input channels might be less efficient than combining their latent representations obtained at the end of the encoders.

Chapter 5

Results and Analysis

In this chapter, we will describe the different results obtained and the corresponding analysis, also visualizing some of the examples produced by different models.

The objective, that was described in the previous chapter, is to compare the performance of the U-Net architecture using the three models considered include: using only contrast-enhanced CT images, a model utilizing both contrast and non-contrast images with two input channels, and a model with two separate encoders for contrast and non-contrast images, followed by a shared decoder. The objective of these models is to determine the impact of adding additional information on the accuracy and detail of the reconstructed images.

The loss function used to train all the proposed models was the MSE. The loss was minimized using the Adam with a learning rate of 0.0001. To mitigate the risk of overfitting, Early Stopping was employed with a patience parameter of 30, ensuring the model stopped training once the best performance on the validation set was achieved.

MSE and SSIM evaluations of model performance apply to the validation and test sets. For a more comprehensive evaluation, also computed are measures like mean, standard deviation, minimum, and maximum MSE.

5.1 Results of the U-Net without side information

Table 5.1: Comparison of MSE and SSIM metrics for the U-Net model under different configurations (256, 128, and 64 lines). The table includes minimum, maximum, mean, and standard deviation values for both test and validation datasets.

MSE				
Lines	Min (Test)	Max (Test)	Mean (Test)	Std Dev (Test)
256	0.0554	0.1163	0.0726	0.0100
128	0.0555	0.1163	0.0726	0.0100
64	0.0677	0.1291	0.0927	0.0115
Lines	Min (Val)	Max (Val)	Mean (Val)	Std Dev (Val)
256	0.0599	0.0912	0.0715	0.0059
128	0.0599	0.0912	0.0715	0.0059
64	0.0661	0.1164	0.0913	0.0100

SSIM				
Lines	Min (Test)	Max (Test)	Mean (Test)	Std Dev (Test)
256	0.2293	0.6137	0.3957	0.0841
128	0.2019	0.6032	0.3612	0.0931
64	0.1821	0.3525	0.2076	0.0251
Lines	Min (Val)	Max (Val)	Mean (Val)	Std Dev (Val)
256	0.2694	0.5618	0.3863	0.0658
128	0.2231	0.5446	0.3449	0.0740
64	0.1853	0.2563	0.2026	0.0148

In Table 5.1, the 256-line configuration was trained for 28 epochs, the 128-line configuration was trained for 114 epochs and the 64-line configuration was trained for 183 epochs. With these values, it might be to verify that the more lines the better the model performance is improved. It is also easy to see that the 256-line configuration achieved better results with the lowest minimum and maximum values for both MSE, indicating a better reconstruction quality. On the contrary, the 64-line configuration achieved the worst results as it had the highest values in the MSE metric, suggesting a lower reconstruction quality and greater variability.

To better understand the results obtained in table 5.1 it's necessary to compare these metrics with the real images having the filtered projection used as input to the network. The results showed that, for the filtered projection, the average MSE with 256 lines was 0.0834, with 128 lines was 0.1002 and with 64 lines was 0.1271 and the SSIM with 256 lines was 0.4193, with 128 lines was 0.3812 and with 64 lines was 0.1984. By comparing the values, it was verified that using the U-Net network, it managed to clean the images, significantly reducing the MSE and improving the SSIM in all tested configurations, thus having a positive impact on the reconstruction of the images.

As expected, when the number of lines decreases from 256 to 64, there is a lower quality of image reconstruction.

In Table 5.1, it's observed that both the 256 rows and the 128 rows achieved the same maximum values, averages, and standard deviations in terms of MSE during training the U-Net model, the main difference between the two is the number and times that each configuration took to reach these values. With these results, it can be concluded that to achieve the best possible performance in a faster and more effective way it would have to be used the greatest number of lines.

Some similarities and differences can be seen regarding the results obtained by the U-Net model with different numbers of lines (256, 128, and 64) when we make a comparison between the Test SSIM Table and the Validation SSIM Table in Table 5.1:

The SSIM values in the test set tend to be higher than those in the validation set, which indicates that the reconstruction quality is slightly higher in terms of structural similarity, as the minimum, maximum, and average SSIM values are all higher in the test set. Regarding standard deviation, the 256-row configuration shows better performance as the test set has a smaller value, which suggests that the results are more stable.

We can therefore conclude from this analysis that the test set can generate more reliable reconstructions, especially when using a greater number of lines, as evidenced by the better SSIM metrics and lower variability.

Looking into the results it is possible to see that the SSIM metrics are superior in the test set compared to the validation set. It is also possible to see that the 256-row configuration gives the best results in both MSE and SSIM metrics.

MSE and SSIM Showing the 3 Best vs the 3 Worst Images:

Test Set: 256 Lines

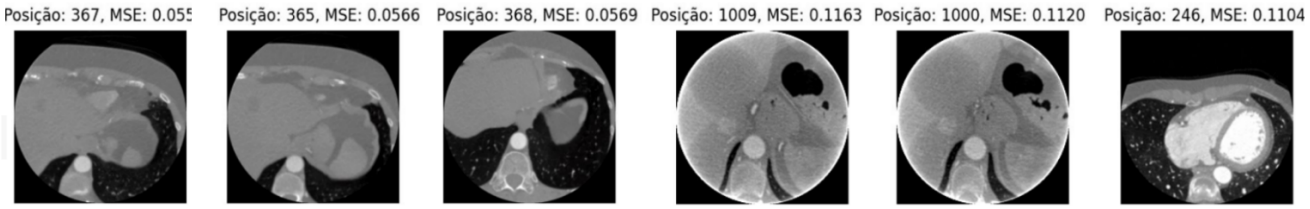


Figure 5.1: U-NET Images with the 3 best (left) and the 3 worst (right) MSEs (256 lines, Test Set).

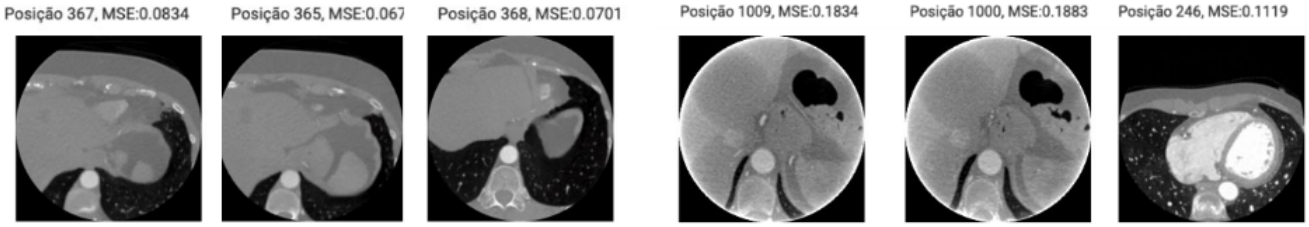


Figure 5.2: Corresponding FBP Images related to the ones above (256 lines, Test Set).

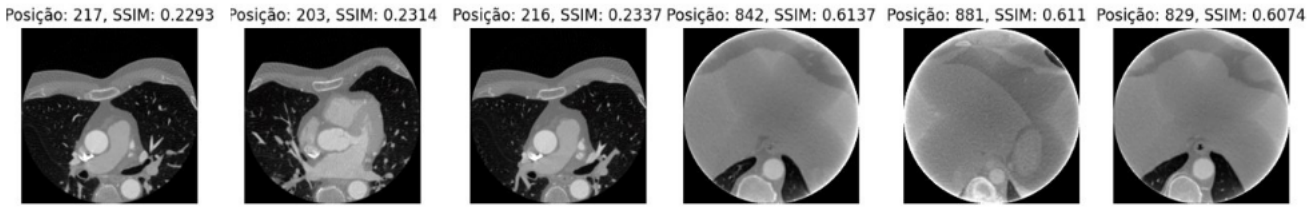


Figure 5.3: U-NET Images with the 3 best (left) and the 3 worst (right) SSIMs (256 lines, Test Set).

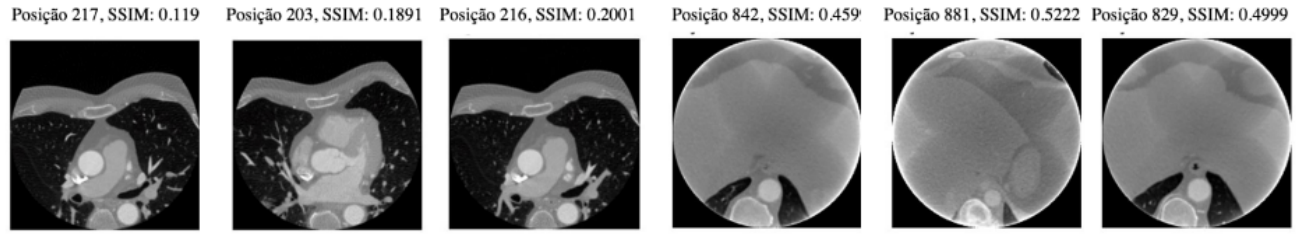


Figure 5.4: Corresponding FBP Images related to the ones above (256 lines, Test Set).

The U-Net network model, when using 256 lines, can present good image quality when reconstructing high-contrast and well-defined images, however, when the images have low contrast, where the structures are thin or complex or with many artifacts (for example blurring, jagged edges, or noise.), the image, when rebuilt no longer has as much quality. When there are many artifacts, it is because the images have higher MSE values and lower SSIM values, as the model has more difficulty reconstructing the images, leading to blurred or distorted images in the output image.

Test Set: 128 Lines

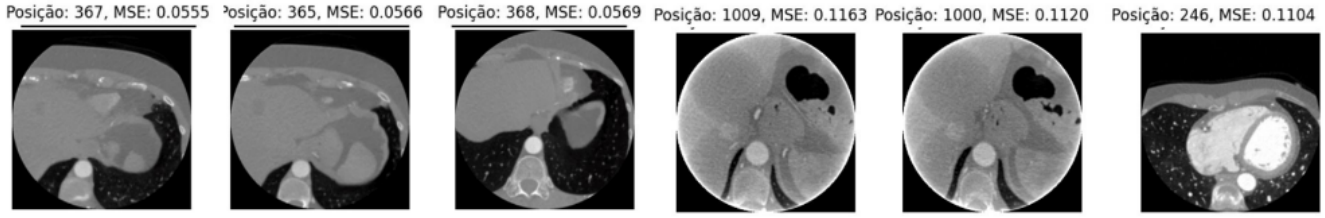


Figure 5.5: U-NET Images with the 3 best (left) and the 3 worst (right) MSEs (128 lines, Test Set).

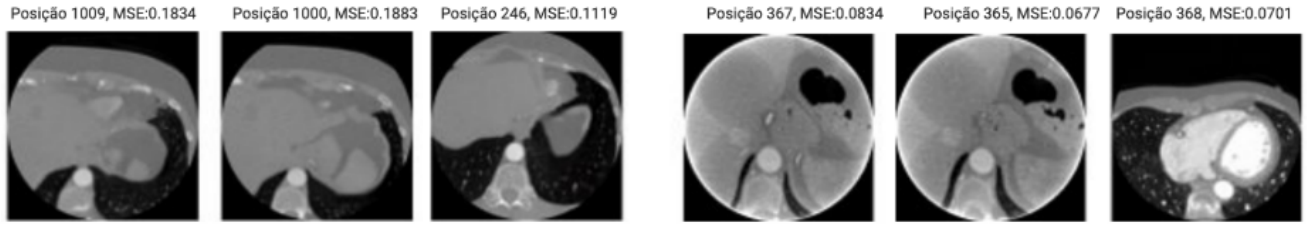


Figure 5.6: Corresponding FBP Images related to the ones above (128 lines, Test Set).

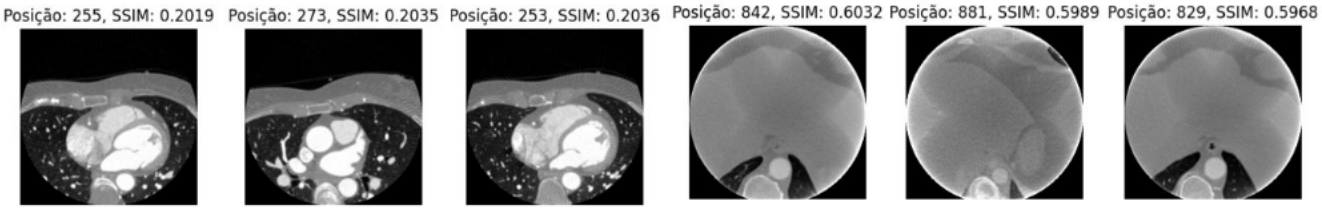


Figure 5.7: U-NET Images with the 3 best (left) and the 3 worst (right) SSIMs (128 lines, Test Set).

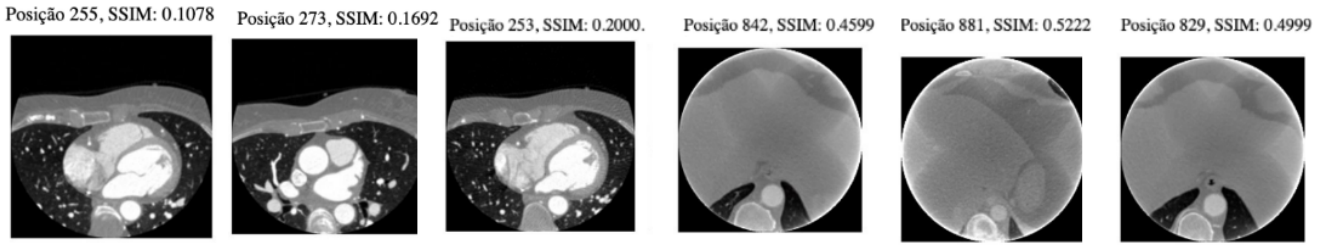


Figure 5.8: Corresponding FBP Images related to the ones above (128 lines, Test Set).

The U-Net network model, when using 128 lines, shows a decrease in performance compared to that of 256 lines, but those are not very relevant.

Test Set: 64 Lines

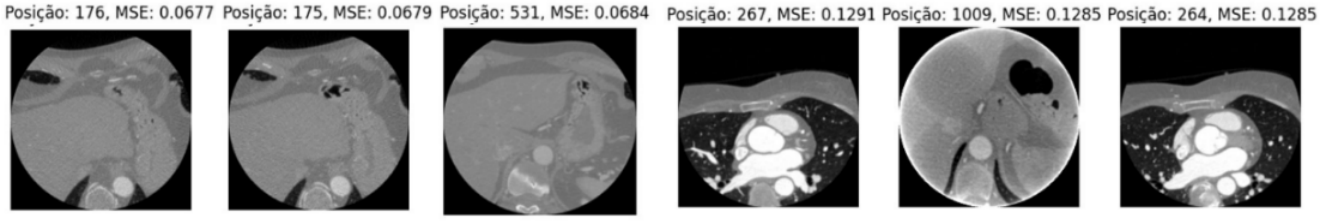


Figure 5.9: U-NET Images with the 3 best (left) and the 3 worst (right) MSEs (64 lines, Test Set).

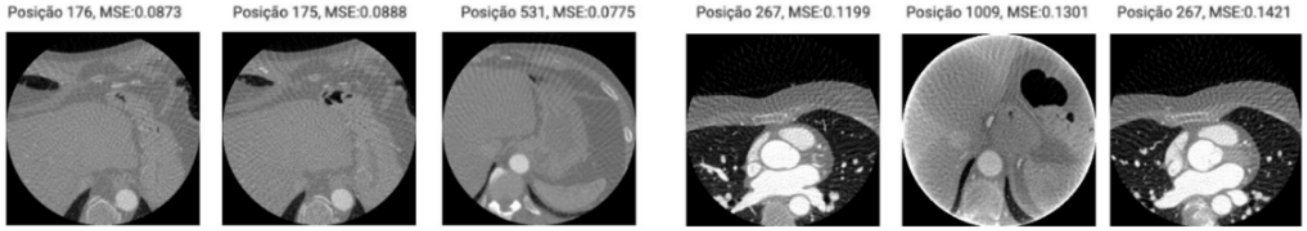


Figure 5.10: Corresponding FBP Images related to the ones above (64 lines, Test Set).

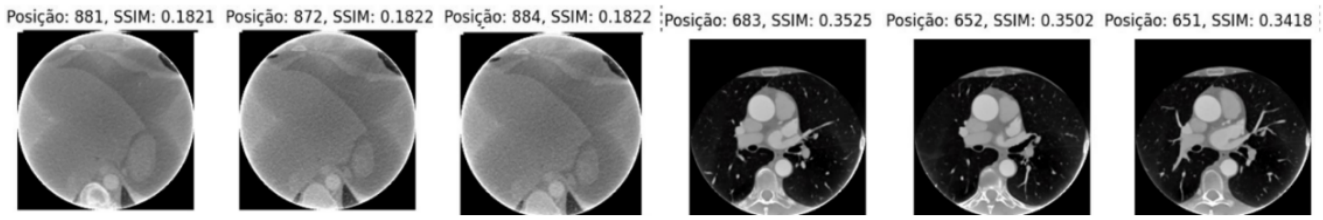


Figure 5.11: U-NET Images with the 3 best (left) and the 3 worst (right) SSIMs (64 lines, Test Set).

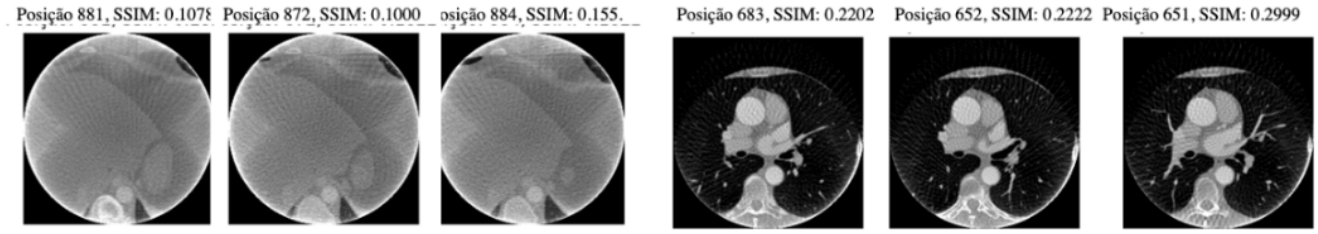


Figure 5.12: Corresponding FBP Images related to the ones above (64 lines, Test Set).

The U-Net network model, when using 64 lines, already has a large difference in quality compared to that of 128 lines, as the image reconstruction loses details, since the noise in the images does not have a large reduction. The model has more difficulty reconstructing the image, which is reflected in higher MSE values and lower SSIM values.

5.2 Results of the U-Net with two input channels

Table 5.2: Comparison of MSE and SSIM metrics for the U-Net model under different configurations (256, 128, and 64 lines). The table includes minimum, maximum, mean, and standard deviation values for both test and validation datasets.

MSE				
Lines	Min (Test)	Max (Test)	Mean (Test)	Std Dev (Test)
256	0.0312	0.0823	0.0601	0.0100
128	0.0389	0.0900	0.0672	0.0125
64	0.0556	0.1073	0.0887	0.0109
Lines	Min (Val)	Max (Val)	Mean (Val)	Std Dev (Val)
256	0.0368	0.0702	0.0566	0.0082
128	0.0418	0.0699	0.0528	0.0070
64	0.0521	0.0973	0.0744	0.0100

SSIM				
Lines	Min (Test)	Max (Test)	Mean (Test)	Std Dev (Test)
256	0.4796	0.8956	0.7211	0.0611
128	0.4032	0.6998	0.5412	0.0798
64	0.2573	0.4162	0.4001	0.0341
Lines	Min (Val)	Max (Val)	Mean (Val)	Std Dev (Val)
256	0.4829	0.8373	0.6123	0.0512
128	0.3913	0.6832	0.4812	0.0683
64	0.2603	0.3819	0.3512	0.0122

In Table 5.2, the 256-line configuration was trained for 40 epochs, the 128-line configuration was trained for 119 epochs and the 64-line configuration was trained for 215 epochs. The configuration with 256 lines achieved greater consistency and better results in terms of MSE, SSIM, and standard deviation, while the configuration with 128 and 64 lines followed a decreasing performance trend, with increases in MSE and decreases in SSIM values as the number of lines decreased. Therefore is possible to conclude that the 256 lines are better in terms of accuracy and consistency of reconstructions.

In order to have a deeper comprehension of the outcomes shown in Table 5.2, it is necessary to juxtapose these metrics with the actual image alongside the filtered projection that was used as input for the network. The findings indicate that, when using the filtered projection, the MSE was 0.0834 for 256 lines, 0.1002 for 128 lines, and 0.1271 for 64 lines. Furthermore, for 256 lines the SSIM was 0.4193 for 128 lines it was 0.3812 and for 64 lines it was 0.1984. By significantly lowering the MSE and improving the SSIM in all investigated configurations, it can be shown by value comparison that the deployment of the U-Net network effectively improves image quality. As such, this improved the process of image reconstruction.

As it's possible to see in Table 5.2, if the SSIM values obtained in the testing are compared with the validation phases, it's verified the differences and similarities in the performance of the U-Net model for different line configurations (256, 128, and 64). As in the previous model, both the minimum SSIM values in the test phase are higher than the values in the validation phase, and the maximum SSIM values in the test phase are higher than those in the validation phase. At last, we may also note a little discrepancy in the test findings shown by a smaller standard deviation, which indicates a higher consistency than in the validation data.

In summary, the results indicate a trend toward higher SSIM metrics in the test set compared to the validation set, signaling slightly better performance in test-based reconstructions regarding to structural fidelity. The 256-row configuration consistently achieves superior levels of performance in both contexts.

MSE and SSIM Showing the 3 Best vs the 3 Worst Images:

Test Set: 256 Lines

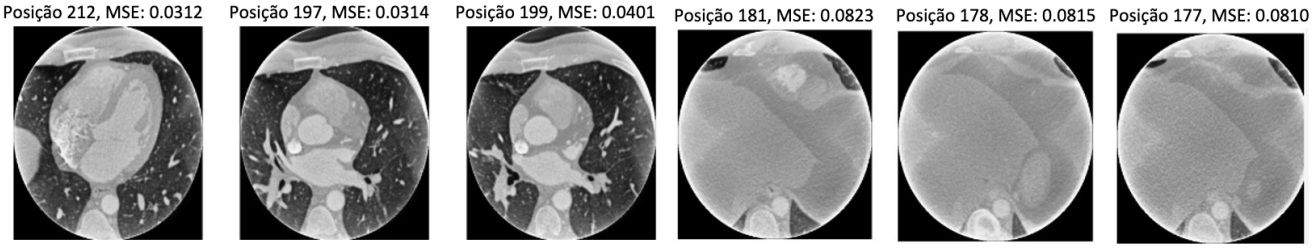


Figure 5.13: U-NET Images with the 3 best (left) and the 3 worst (right) MSEs (256 lines, Test Set).

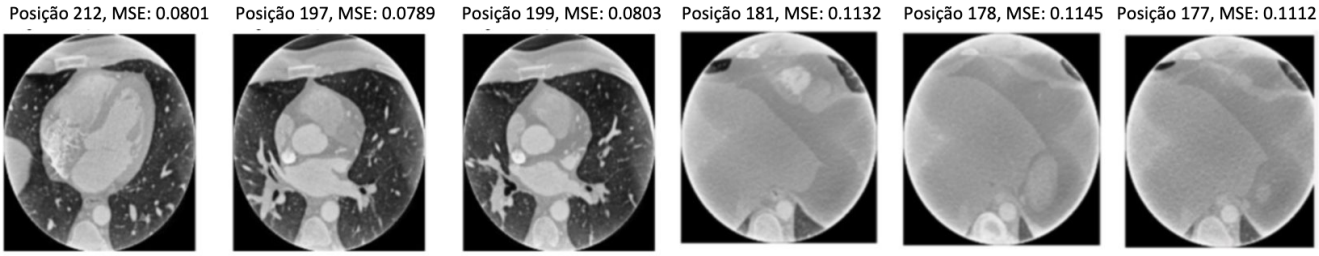


Figure 5.14: FBP Images corresponding to the U-NET images above (256 lines, Test Set).

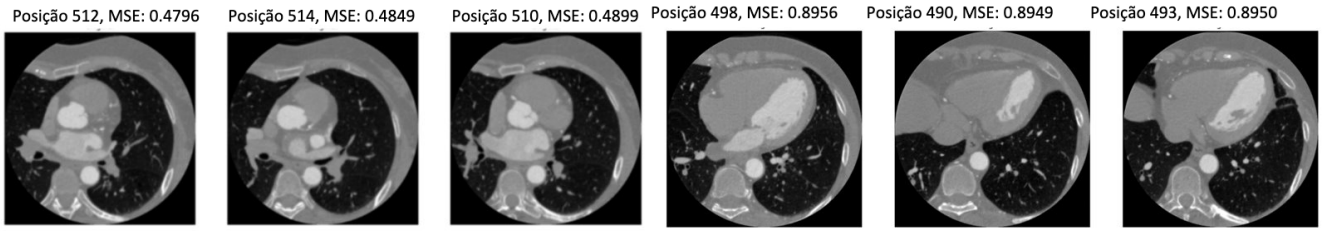


Figure 5.15: U-NET Images with the 3 best (left) and the 3 worst (right) SSIMs (256 lines, Test Set).

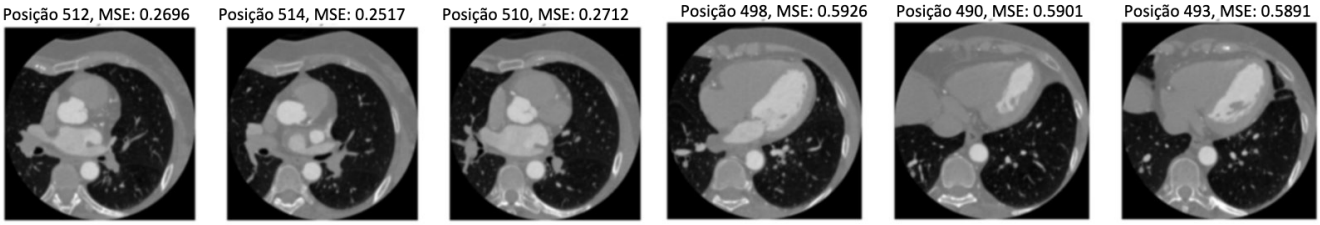


Figure 5.16: FBP Images corresponding to the U-NET images above (256 lines, Test Set).

The U-Net produces excellent reconstructions for 256 lines with the best MSE images similar to the original, the poorest MSE images display obvious blurring and streaking errors. Regarding SSIM, whilst the poorest images expose notable distortions, the best images maintain a high structural closeness to the original. Compared to U-Net outputs, the FBP images for both MSE and SSIM examples display less detail and more obvious artifacts.

Test Set: 128 Lines

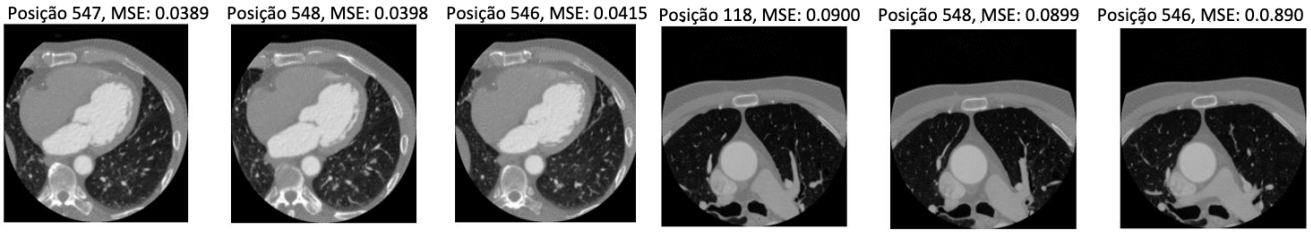


Figure 5.17: U-NET Images with the 3 best (left) and the 3 worst (right) MSEs (128 lines, Test Set).

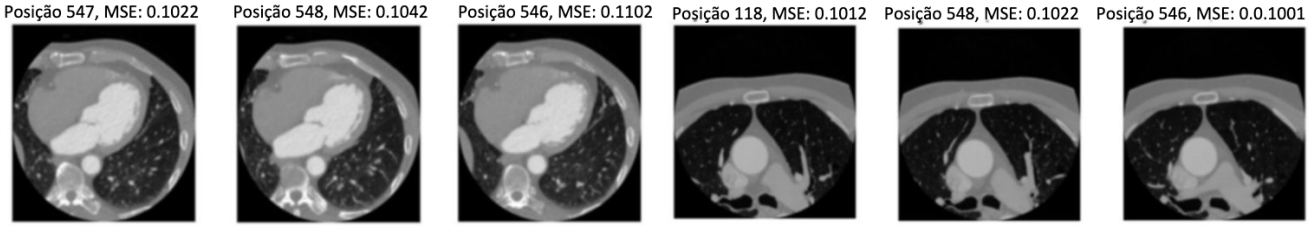


Figure 5.18: FBP Images corresponding to the U-NET images above (128 lines, Test Set).

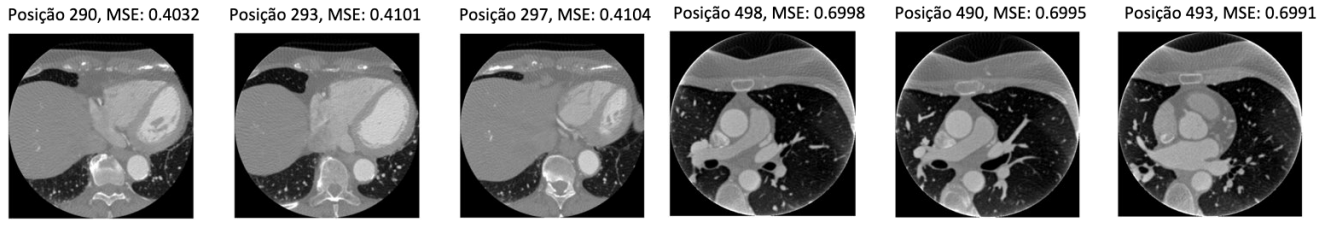


Figure 5.19: U-NET Images with the 3 best (left) and the 3 worst (right) SSIMs (128 lines, Test Set).

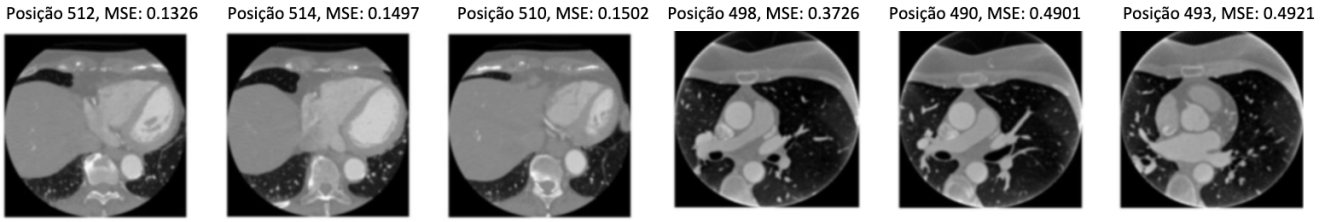


Figure 5.20: FBP Images corresponding to the U-NET images above (128 lines, Test Set).

For 128 lines, compared to 256 lines, they demonstrate a reduction in quality. The top 3 of the best MSE images show acceptable images but with some artifacts, while the top 3 of the worst images show an increase in blur and artifacts. Regarding SSIM, image quality suffers. We may therefore draw the conclusion that the images acquired by Facebook P display less detail and more obvious artifacts as the number of lines decreases.

Test Set: 64 Lines

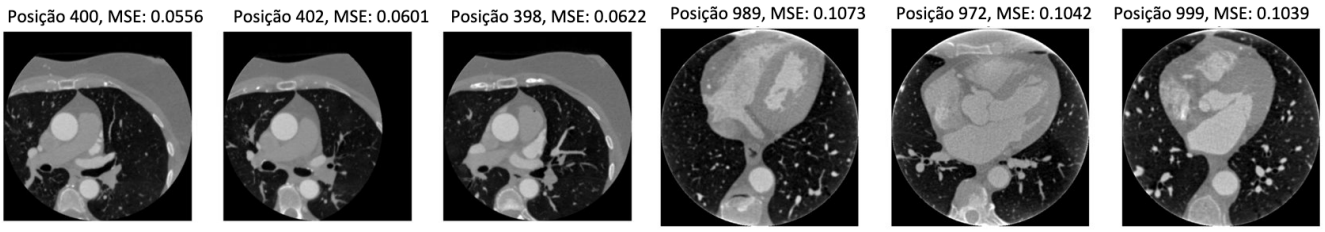


Figure 5.21: U-NET Images with the 3 best (left) and the 3 worst (right) MSEs (64 lines, Test Set).

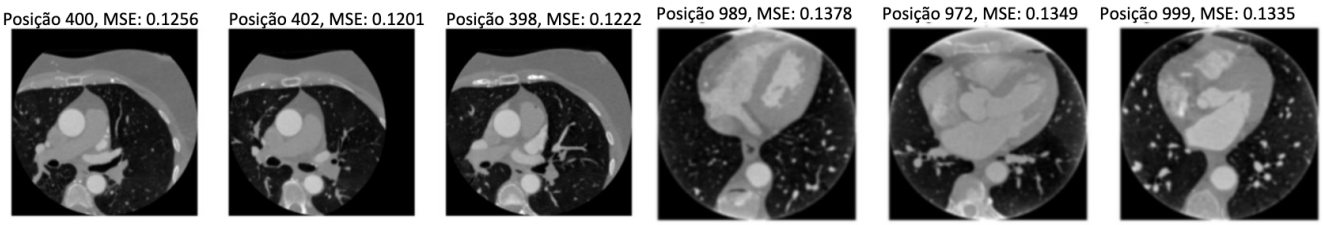


Figure 5.22: FBP Images corresponding to the U-NET images above (64 lines, Test Set).

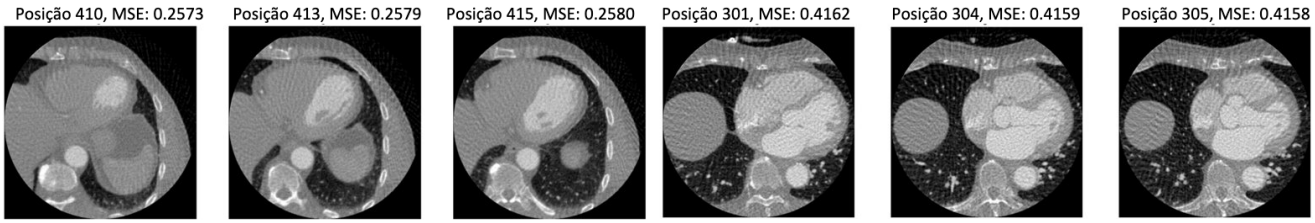


Figure 5.23: U-NET Images with the 3 best (left) and the 3 worst (right) SSIMs (64 lines, Test Set).

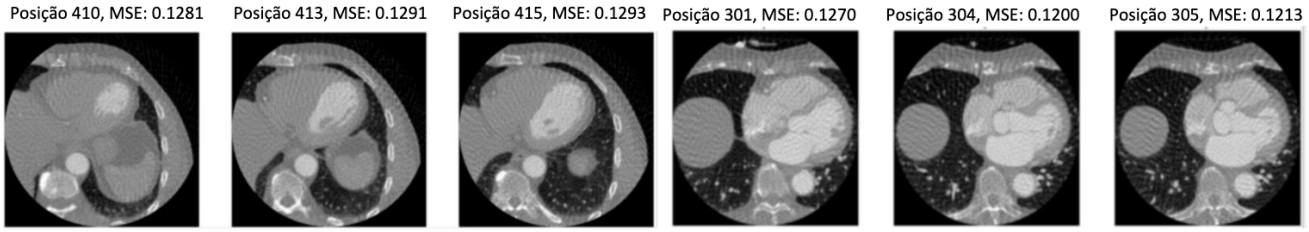


Figure 5.24: FBP Images corresponding to the U-NET images above (64 lines, Test Set).

For 64 lines, U-Net performance is substantially degraded, even the best MSE images show a lot of artifacts, and the worst images are heavily blurred and have severe artifacts. Relative to SSIM, they also show a significant loss of structural detail. The images for 64 lines are very noisy and exhibit extreme artifacts.

5.3 Results of the U-Net with two encoders

Table 5.3: Comparison of MSE and SSIM metrics for the U-Net model under different configurations (256, 128, and 64 lines). The table includes minimum, maximum, mean, and standard deviation values for both test and validation datasets.

MSE				
Lines	Min (Test)	Max (Test)	Mean (Test)	Std Dev (Test)
256	0.0215	0.0592	0.0399	0.0100
128	0.0285	0.0618	0.0425	0.0105
64	0.0331	0.0711	0.0503	0.0114
Lines	Min (Val)	Max (Val)	Mean (Val)	Std Dev (Val)
256	0.0255	0.0512	0.0419	0.0061
128	0.0301	0.0603	0.0499	0.0069
64	0.0313	0.0841	0.0581	0.0102

SSIM				
Lines	Min (Test)	Max (Test)	Mean (Test)	Std Dev (Test)
256	0.6923	0.9009	0.7919	0.0699
128	0.6033	0.8001	0.6832	0.0692
64	0.4923	0.6821	0.5813	0.0399
Lines	Min (Val)	Max (Val)	Mean (Val)	Std Dev (Val)
256	0.6823	0.9001	0.8087	0.0401
128	0.4918	0.8771	0.7014	0.0523
64	0.4001	0.6242	0.5161	0.1001

Compared with the other 2 previous models it's possible to visualize an improvement in MSE and SSIM.

Observing table 5.3, regarding to the 256 lines, it shows the lowest average MSE and the highest average SSIM, compared to 128 lines and 64 lines there was a noticeable decline in relation to the MSE and SSIM values.

To fully understand the results shown in Table 5.3, it is essential to compare these metrics with the real images and the filtered projection that was used as input for the network. The results suggest that, when using the filtered projection, the MSE was 0.0834 for 256 lines, 0.1002 for 128 lines, and 0.1271 for 64 lines. The SSIM values were 0.4193 for 256 lines, 0.3812 for 128 lines, and 0.1984 for 64 lines. By conducting a value comparison, it is evident that the use of the U-Net network effectively improved the quality of the images by significantly reducing the MSE and boosting the SSIM) in all the tested configurations. As a result, this had a beneficial impact on the process of reconstructing the image.

As observed in the two previous models, the data suggests a strong correlation between the number of lines and the quality of the reconstruction, so the greater the number of lines, the better the image quality and the MSE.

The comparison between SSIM Test and SSIM Validation in Table 5.3 reveals the differences and similarities in the results obtained by the U-Net model:

Regarding the results obtained for the SSIM metric, the minimum test values are higher compared to the minimum validation values, which means better accuracy in image reconstruction. Regarding the maximum values of SSIM, it's observed better results in the test set than in the validation set, which also indicates a higher degree of accuracy in image reconstructions. In relation to the average SSIM values, they also have higher averages in the test set.

Concluding, in general, the third model is the best performing, more specifically with 256 lines, it shows better average performance in both SSIM and MSE.

MSE and SSIM Showing the 3 Best vs the 3 Worst Images:

Test Set: 256 Lines

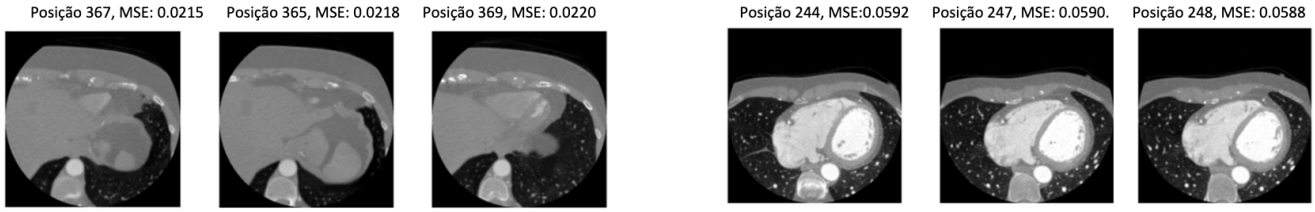


Figure 5.25: U-NET Images with the 3 best (left) and the 3 worst (right) MSEs (256 lines, Test Set).

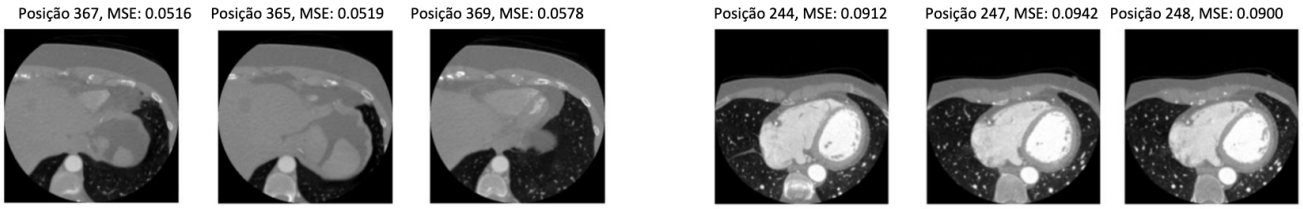


Figure 5.26: FBP Images related with the image above.

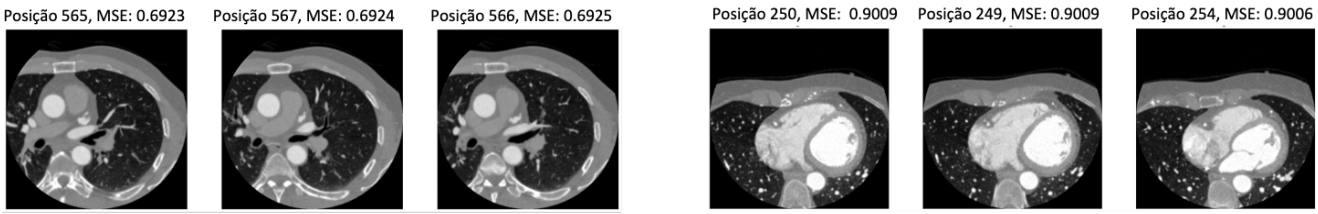


Figure 5.27: U-NET Images with the 3 best (left) and the 3 worst (right) SSIMs (256 lines, Test Set).

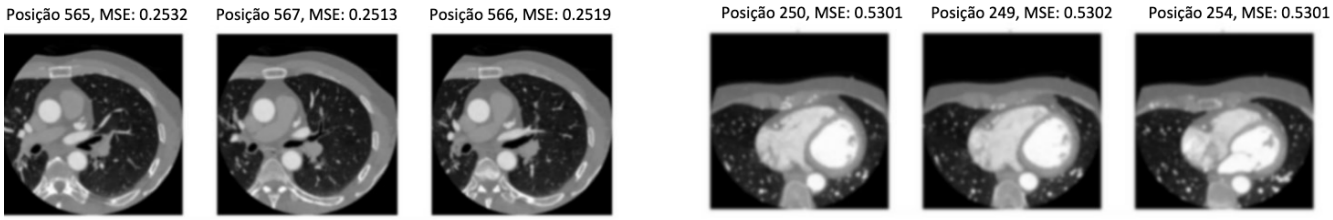


Figure 5.28: FBP Images related with the image above.

For 256 lines, the U-Net network demonstrates a higher quality of image reconstruction, with the best images from MSE and SSIM being very close to the original ones, while the worst ones show only a few artifacts, I can conclude that the Images demonstrate more detail and fewer artifacts.

Test Set: 128 Lines

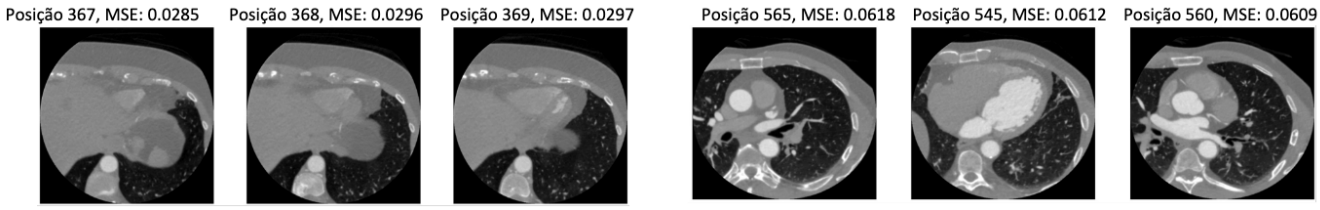


Figure 5.29: U-NET Images with the 3 best (left) and the 3 worst (right) MSEs (128 lines, Test Set).

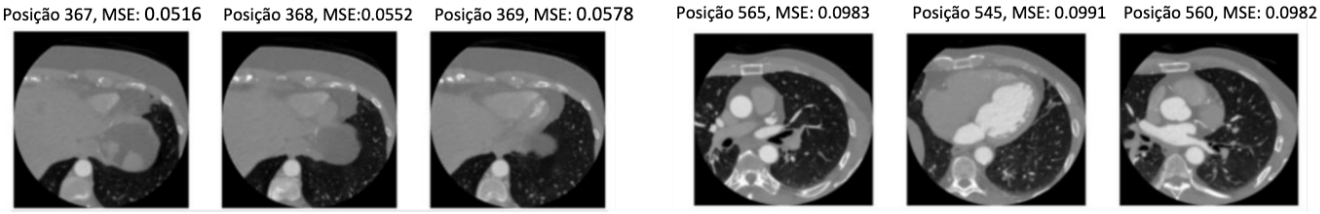


Figure 5.30: FBP Images related with the image above.

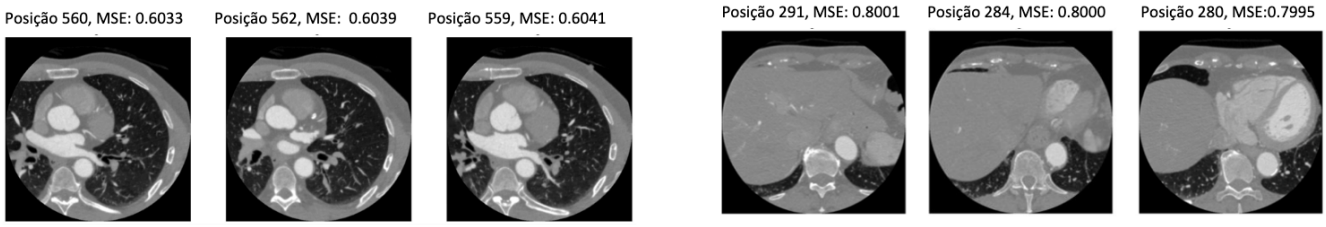


Figure 5.31: U-NET Images with the 3 best (left) and the 3 worst (right) SSIMs (128 lines, Test Set).

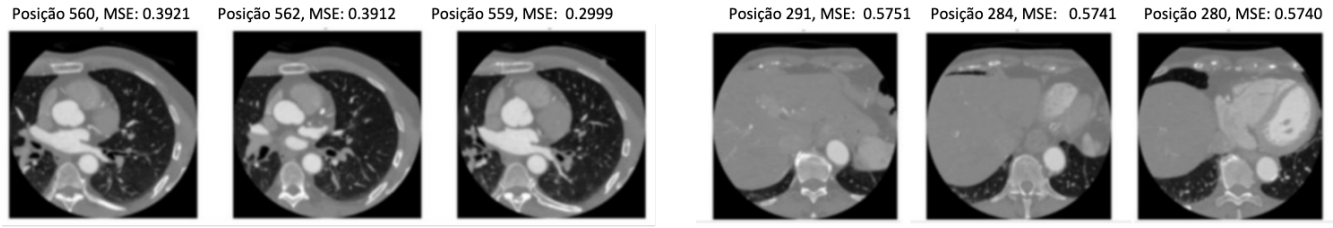


Figure 5.32: FBP Images related with the image above.

With 128 lines, compared to 256 lines, there is a reduction in quality, but the best images from both MSE and SSIM are clean images with good quality, but the worst images already show more artifacts and blurs.

Test Set: 64 Lines

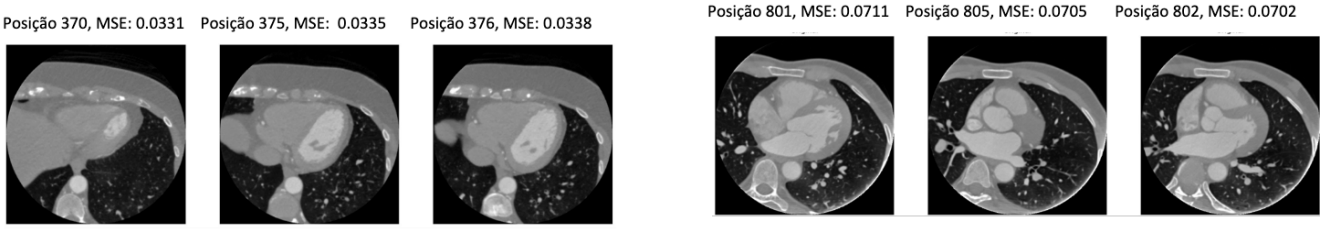


Figure 5.33: U-NET Images with the 3 best (left) and the 3 worst (right) MSEs (64 lines, Test Set).

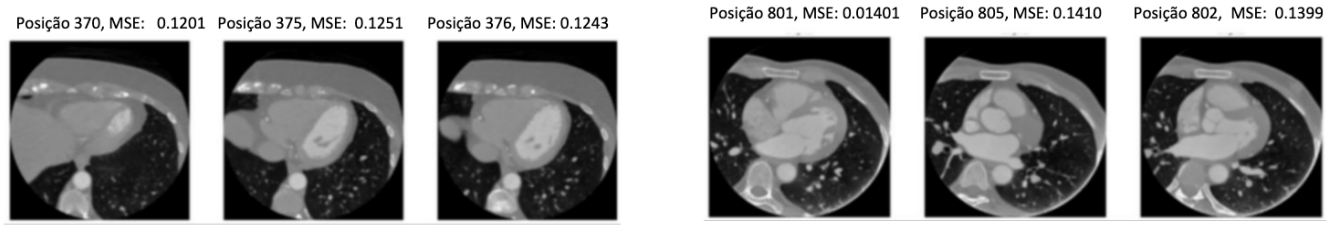


Figure 5.34: FBP Images related with the image above.

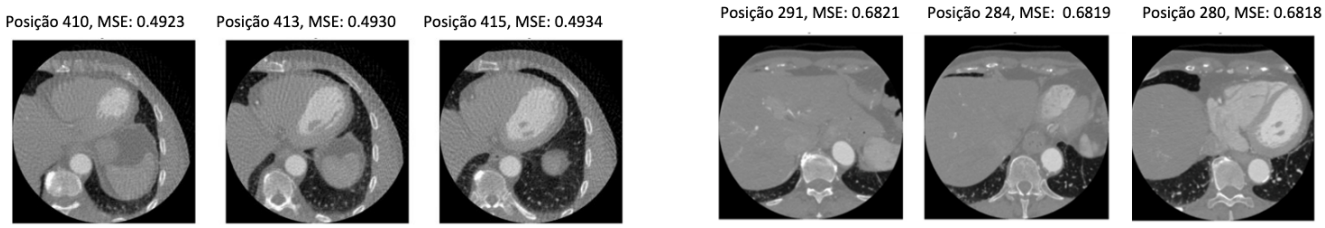


Figure 5.35: U-NET Images with the 3 best (left) and the 3 worst (right) SSIMs (64 lines, Test Set).

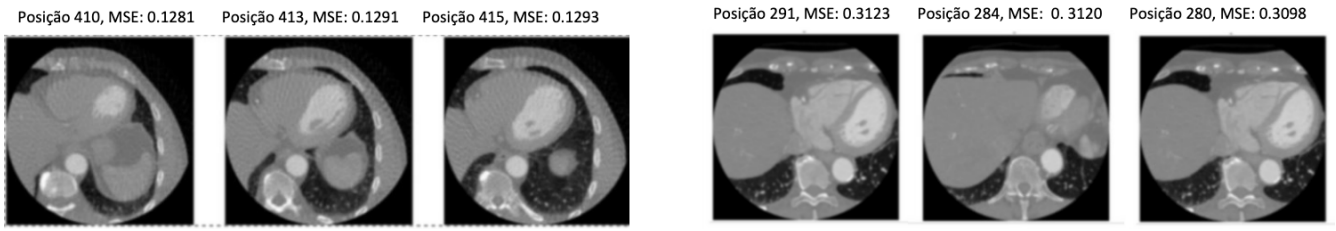


Figure 5.36: FBP Images related with the image above.

For the 64 lines, if it's compared to the previous models, it lost immense quality in image reconstruction, suffered blurring, loss of details, and increased noise, especially in the worst cases. When images are of low contrast, it is very difficult to maintain structural integrity, resulting in higher MSE and lower SSIM values.

5.4 Comparing image with the 3 different models

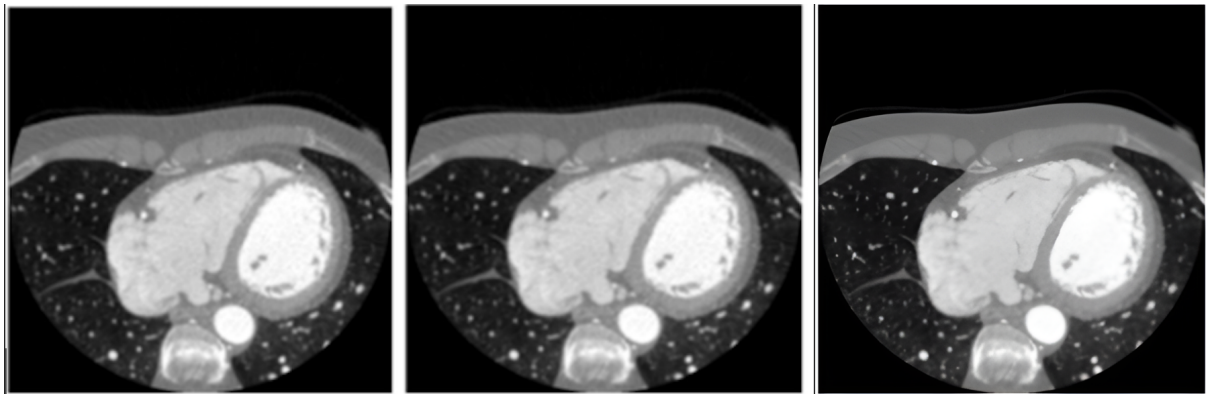


Figure 5.37: image without side information (left), image with side information using 2 channels (center) and image with side information using 2 encoders (right)

5.5 Discussion

By obtaining the results from these images, it was possible to collect the following information: images with 64 lines show more obvious artifacts, for example blurring and loss of clarity in fine structures, while images reconstructed with 256 lines show fewer visual artifacts than those reconstructed with 128 lines. Reconstruction of finer and more complex structures clearly becomes challenging, especially as the number of lines decreases. Particularly with few lines (64 and 128), the U-Net model sometimes loses detail in these complex areas. On the other hand, in homogeneous areas, the reconstruction is more robust, even in images with 64 lines. However, accuracy decreases at edges and transitions, where there is a variation in intensity. In general, images reconstructed with 256 lines are superior to those reconstructed with 128 and 64 lines this suggests that more reliable reconstructions come from greater data availability.

By using non-contrast CT scans as additional information, the comparison between these models shows the influence of this information on the reconstruction of contrast-enhanced CT scans from low-dose data. It can be observed that the quality of the reconstruction was lower in models without additional information, particularly in areas that require detailed preservation of the structure. In contrast-enhanced images with 256 lines, models that used additional information showed a significant improvement in preserving fine details and reducing artifacts. Even with fewer lines, the additional data provided more context, which improved the model's ability to accurately reconstruct the image.

In summary, it can be seen that among the three models tested, the U-Net design with two encoders that uses additional information proved to be the most efficient. By efficiently utilizing the additional non-contrast data, this model outperformed the others, especially in situations requiring low-dose measurements. This side-by-side comparison of the models illustrates the

significant advantage of including additional information, especially in improving reconstruction quality and preserving structural features across different line configurations.

Chapter 6

Conclusion

This dissertation aimed to utilize low-dose CT images and use sophisticated DL approaches to enhance the clarity and quality of images with patients suffering from CAD. As discussed in the thesis, coronary disease is a serious global health issue, however proper diagnosis depends on a clean and quality computed tomography (CT) image so that physicians may provide efficient therapy. Although contrast-enhanced CT is the most often used everyday method, it has disadvantages including the need for significant radiation doses.

With a goal to lower radiation exposure, the dissertation investigated the U-Net architecture for the reconstruction of low-dose CT images, thereby enabling either maintenance or even improvement of the reconstruction images, attaining image quality and accuracy. Three distinct techniques were applied: reconstruction without side information, reconstruction with side information using two channels, and without side information using two encoders utilizing images with contrast and without contrast.

As it is feasible to examine the MSE and SSIM indicators, the findings revealed that incorporating extra information via two encoder channels might significantly increase the quality of the reconstructed images. The U-Net model that used 256 lines as input had better results than those that used fewer lines, therefore demonstrating that the higher the data density, the better the reconstruction quality and accuracy and structural fidelity.

We therefore draw the conclusion that, with respect to low-dose CT imaging for the diagnosis of CAD, the approaches underlined throughout the dissertation show promise, especially the U-Net design with 2 encoders. These techniques provided a dependable and quick diagnostic tool in addition to helping patients's radiation dose be lowered.

For future work, they will be able to concentrate on enhancing these models and investigating their use in clinical environments to raise patient outcomes for next projects.

Bibliography

[1] Baystate Health. History of heart disease. 2024.

[2] Slicer community. 3d slicer. https://www.slicer.org/, accessed March 12, 2023.

[3] Maurizio Conti, Peter Valk, Sherif Heider, Lucas Ferrer, Valeria Racca, and Madan M. Rehani. Artificial intelligence and machine learning: Relevance for the optimization of pet and spect imaging in clinical practice. *European Journal of Nuclear Medicine and Molecular Imaging*, 50(3):531–547, 2023. doi:10.1007/s00259-022-05746-4.

[4] M. Doris and D.E. Newby. Coronary ct angiography as a diagnostic and prognostic tool: Perspectives from the scot-heart trial. *Current Cardiology Reports*, 18(18), 2016. doi:10.1007/s11886-015-0695-4.

[5] Jawook Gu, Tae Seong Yang, Jong Chul Ye, and Dong Hyun Yang. Cyclegan denoising of extreme low-dose cardiac ct using wavelet-assisted noise disentanglement. *Medical Image Analysis*, 74:102209, 2021. ISSN: 1361-8415. doi:https://doi.org/10.1016/j.media.2021.102209.

[6] Jiuxiang Gu, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, Xingxing Wang, Gang Wang, Jianfei Cai, and Tsuhan Chen. Recent advances in convolutional neural networks. *Pattern Recognition*, 77:354–377, 2018. ISSN: 0031-3203. doi:https://doi.org/10.1016/j.patcog.2017.10.013.

[7] J. Gurrola-Ramos, O. Dalmau, and T.E. Alarcón. A residual dense u-net neural network for image denoising. *IEEE Access*, 9:31742–31754, 2021. doi:10.1109/ACCESS.2021.3059657.

[8] Matej Habijan, Hrvoje Leventić, Ivana Galić, and Danilo Babin. Whole heart segmentation from ct images using 3d u-net architecture. In *2019 International Conference on Systems, Signals and Image Processing (IWSSIP)*, pages 121–126, Osijek, Croatia, 2019. doi:10.1109/IWSSIP.2019.8787253.

[9] Mattias P. Heinrich, Maik Stille, and Thorsten M. Buzug. Residual u-net convolutional neural network architecture for low-dose ct denoising. *Current Directions in Biomedical Engineering*, 4(1):297–300, 2018. doi:10.1515/cdbme-2018-0072.

[10] Yang Jing and Yang Guanci. Modified convolutional neural network based on dropout and the stochastic gradient descent optimizer. *Algorithms*, 11:28, 03 2018. doi:10.3390/a11030028.

[11] Eunju Kang, Jinkee Min, and Jong Chul Ye. A deep convolutional neural network using directional wavelets for low-dose x-ray ct reconstruction. *Med. Phys.*, 44:e360–e375, 2017. doi:10.1002/mp.12344.

[12] K. Lee and H. Kim. Advanced denoising techniques for ct images using convolutional neural networks. *IEEE Transactions on Medical Imaging*, 40(2):556–567, 2021.

[13] Hualin Liu, Zhe Li, Shijie Lin, and Libo Cheng. A residual u-net denoising network based on multi-scale feature extraction and attention-guided filter. *IEEE Access*, 9:56789–56799, 2021. doi:10.1109/ACCESS.2021.3067890.

[14] Johns Hopkins Medicine. Anatomy and function of the coronary arteries, Accessed 2024.

[15] Mount Sinai Health System. Hardening of the arteries. https://www.mountsinai.org/health-library/diseases-conditions/hardening-of-the-arteries.

[16] Author(s) Name(s). Convolutional neural network: An overview. *Analytics Vidhya*, January 2022.

[17] Karen Okrainec, Devi K Banerjee, and Mark J Eisenberg. Coronary artery disease in the developing world. *American Heart Journal*, 148(1):7–15, 2004. ISSN: 0002-8703. doi:10.1016/j.ahj.2004.03.016.

[18] Prajeesh Prathap. The secret to understanding cnns: Convolution, feature maps, pooling, and fully connected layers. *Medium*, Month Year.

[19] U Ralapanawa and R Sivakanesan. Epidemiology and the magnitude of coronary artery disease and acute coronary syndrome: A narrative review. *Journal of Epidemiology and Global Health*, 11(2):169–177, Jun 2021. doi:10.2991/jegh.k.201217.001.

[20] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William Wells, and Alejandro Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015*, volume 9351 of *Lecture Notes in Computer Science*, pages 234–241. Springer, Cham, 2015. doi:10.1007/978-3-319-24574-4$_2$8.

[21] J. Smith and A. Doe. Deep learning for noise reduction in medical imaging. *International Journal of Computer Vision*, 128(4):987–1001, 2020.

[22] M.M. Taye. Understanding of machine learning with deep learning: Architectures, workflow, applications and future directions. *Computers*, 12:91, 2023. doi:10.3390/computers12050091.

[23] M.J. Willemink and P.B. Noël. The evolution of image reconstruction for ct—from filtered back projection to artificial intelligence. *European Radiology*, 29:2185–2195, 2019. doi:10.1007/s00330-018-5810-7.

[24] Jelmer M. Wolterink, Bob D. de Vos, Tim Leiner, Max A. Viergever, and Ivana Išgum. Orca score. https://orcascore.grand-challenge.org/, January 2022.

[25] Xun Yi and Paul Babyn. Sharpness-aware low-dose ct denoising using conditional generative adversarial network. *Journal of Digital Imaging*, 31(5):655–669, 2018. doi:10.1007/s10278-018-0056-0.

[26] Y. Zhang, H. Yu, and G. Wang. Du-gan: Generative adversarial networks with dual-domain u-net based discriminators for low-dose ct denoising. *arXiv preprint arXiv:2108.10772*, 2021.

[27] B. Zhou, F. Gao, and Y. Zhang. A deep learning-based image reconstruction method for electrical capacitance tomography. *Signal, Image and Video Processing*, 2023. doi:10.1007/s11760-023-02528-9.