UNIVERSITY OF CAMPINAS

INSTITUTE OF COMPUTING

Master's Degree Proposal

17/10/2024

Enhancing Synchrotron Low-Dose Computed Tomography Image Quality Using Diffusion-Based Generative Models

Candidate: Paulo Baraldi Mausbach

Supervisor: Prof. Dr. Zanoni Dias

Cosupervisor: Prof. Dr. Hélio Pedrini

Contents

Abstract

Medical and synchrotron Computed Tomography (CT) takes advantage of eletromagnetic radiation and its interaction with matter to assess internal structures of their objects of study. Due to radiation dose concerns, Low-Dose Computed Tomography (LDCT) became an essential technique for reducing radiation exposure by the cost of introducing noise and artifacts, challenging specialists to innovate on how to enhance image quality. This research explores the potential of diffusion-based generative models for denoising synchrotron LDCT images. By leveraging existing medical LDCT datasets and repurposing models trained over them for synchrotron LDCT denoising, this study aims to enhance synchrotron LDCT image quality on post-processing stage. The research investigates the effectiveness of diffusion models, compares their performance with other methods, and explores fine-tuning techniques to improve denoising outcomes. The findings are expected to contribute to the development of advanced denoising strategies, providing better image analysis and post-processing capabilities.

Chapter 1

Introduction

In this chapter, the motivations of this work are presented in Section 1.1, the problem definition in Section 1.2, the research questions that guide this project in Section 1.3, the goals and expected results in Section 1.4 and the overall text organization in Section 1.5.

1.1 Motivations

X-ray Computed Tomography (CT) stands as a pivotal noninvasive imaging technique, finding extensive applications across various domains. The CT imaging process can be summarize as capturing 2D transmission projection images of an object from various angles around a common axis, followed by a reconstruction technique to restore the object's 3D morphology [31, 36].

Within medical imaging domain, CT imaging has historically reduced the need for exploratory surgery [9] and is frequently applied clinically as the gold standard for trauma assessment [43]. Beyond medical applications, CT imaging plays an equally significant role in fields like Archaeology [28], Paleontology [23, 44] and Material Science [50], offering a noninvasive way to resolve the internal structure of materials [7], specially when performed on synchrotron facilities as it allows achieving higher spatial resolutions when compared to conventional X-Ray sources [16].

Working with patients or biological samples are often prohibitive due to constraints related to X-Ray radiation dose exposure. On living beings, the exposure to high doses of radiation may be harmful for health [9, 40] while for radiation-sensitive samples it may cause damages to it and directly impact the experiments results [36]. Therefore, researches on Low-Dose Computed Tomography (LDCT) techniques has been done over the decades with the challenge to minimize radiation dose absorbed by objects under analysis without sacrificing image quality [15], usually following the As Low As Reasonably Achievable (ALARA) principle related to radiation exposure [18].

The CT image quality is often defined by two main factors: spatial resolution and contrast resolution [9]. In other words, it means that the ideal tomography image stands on a maximum spatial resolution with the minimum noise. A limitation of LDCT is the trade-off between image noise, patient dose, and z-axis resolution (slice thickness) [9], which inherently results in lower quality images compared to standard CT images [17, 49, 51] and hinder the analysis and post-processing of it. Therefore, the development of denoising strategies are paramount to enhance LDCT image quality and boost post-processing and analysis capabilities over them.

Non-Machine Learning and Machine Learning (ML) approaches have been explored for handling LDCT denoising tasks [42]. However, with the growth of Deep Learning (DL) [32] over the last decade, DL-based solutions have gained prominence [17,31], particularly models utilizing Convolutional Neural Networks (CNN) and Generative Adversarial Network (GAN) architectures, which are currently the most popular approaches [49].

Recent advancements in medical imaging have introduced diffusion-based models as promising tools for a variety of Computer Vision (CV) tasks, including image generation, anomaly detection, reconstruction, registration, and denoising [29]. Denoising Diffusion Probabilistic Models (DDPM) [24] and Stochastic Differential Equations (SDE) [47] have emerged as commonly used methods [29]. However, most studies have not focused specifically on denoising and have typically been conducted on Magnetic Resonance Imaging (MRI) and CT images rather than LDCT images [29], indicating the need for further research on diffusion-based approaches for LDCT denoising.

1.2 Problem Definition

CT images are susceptible to various sources of noise and artifacts, which can compromise image quality. Noise in CT images primarily arises from quantum fluctuations inherent in the photon detection process. Additionally, electronic and structural noise from detection systems and also the reconstruction method chosen and its configuration have their own contribution to the noise in the CT resulting image. Conversely, CT images are also prone to numerous types of artifacts stemming from various sources. Some are due to the polyenergetic nature of the X-Ray spectrum and how it attenuates throughout the sample, while others are related to the reconstruction algorithm chosen, the number of projections used on it or/and the chosen voxel size [6, 9, 42].

The constraints of LDCT related to sample exposure to radiation usually require some balance between experiment time, spatial and contrast resolution. A common approach to reduce radiation dosage is to decrease the number of projections taken and/or the photon flux, with the downside of introducing noise and artifacts to the image due to the lack of data completeness [15].

LDCT denoising approaches are commonly classified in three categories: raw data filtering, iterative reconstruction techniques and post-processing algorithms [36, 42].

Raw data filtering approaches aim to address noise directly in the acquired data by processing it within the sinogram or projection space [17, 42]. While these strategies can effectively reduce noise in LDCT images, they are also reported to cause edge and detail blurring, the appearance of stair effects on edges, and resolution loss [17, 42]. Furthermore, access to raw data is typically restricted due to commercial or research confidentiality concerns [17, 18, 31].

Iterative reconstruction techniques aim to improve reconstruction algorithms, whether model-based or not, to achieve the highest quality images possible [36, 42]. However, these methods are often reported to be computationally expensive [17].

Post-processing algorithms are among the most widely explored approaches, as they do not require access to raw data and bypass the reconstruction step, making them more generally applicable [17, 35]. These methods focus on enhancing the quality of already reconstructed images by applying various image processing techniques [36, 42]. Variants of classical image denoising algorithms, such as Block-Matching and 3D Filtering (BM3D) [13]

and non-local means [8], have been adapted for LDCT denoising, but their performance still requires improvement [17, 56]. CNN-based solutions are the most commonly explored DL approaches and have achieved good performance, but they are often reported to cause over-smoothing effects on the output images and struggle on generalizing for unseen dose levels [17, 18]. Meanwhile, GAN-based approaches aim to handle LDCT denoising tasks while better preserving image details and textures with better generalization. However, their training instability and possibility to generate value shifts are usually considered significant drawbacks [17, 18].

Diffusion-based generative models have shown promising results for denoising tasks in Positron Emission Tomography (PET) [19] and Optical Coherence Tomography (OCT) [26] images, and they have also been applied to LDCT medical images [17, 18, 35, 56], although this area has not been extensively explored yet. Diffusion model approaches have been reported to achieve better denoising and generalization performance over LDCT when compared to CNN and GAN-based methods, with also better stability during training than GANs. Therefore, diffusion models still face challenges with higher inference times due to the need for multiple sampling steps during inference, rather than being end-to-end models [17, 18]. No research has yet been published on applying diffusion models for denoising synchrotron LDCT images, highlighting an opportunity for innovation in this domain.

This study aims to explore the potential of diffusion generative models in addressing denoising tasks in both medical and synchrotron LDCT images while seeking enhancements to current methodologies for improved denoising performance, faster inferences, and domain generalization. Given the scarcity of synchrotron LDCT datasets, repurposing models trained on medical LDCT images is appealing, as both image types share similar noise and artifacts, and medical LDCT datasets [1, 11, 34, 39] are more abundant and larger than synchrotron LDCT datasets [14]. Conversely, self-supervised approaches present an attractive alternative as they do not rely on the existence of extensive labeled datasets for training.

1.3 Research Questions

This work aims to answer the following questions:

- Can diffusion models perform denoising tasks on LDCT reconstructed data toward increasing the quality of synchrotron LDCT?
- Can diffusion models trained over CT medical images be directly repurposed to perform denoising tasks on LDCT synchrotron images?
- Does finetunning a model trained over CT medical images with synchrotron CT images enhances the acquired results?

1.4 Goals and Expected Contributions

The primary objective of this research is to investigate the effectiveness of diffusion-based generative models for perfoming denoising tasks over LDCT from synchrotrons. Secondly, it seeks to assess whether diffusion models trained for denoising medical LDCT images can

generalize well to be applied for denoising synchrotron LDCT images. To achieve these objectives, the following tasks will be undertaken:

- Explore existing solutions that use diffusion-based generative model for denoising medical and synchrotron LDCT images and select the most promising ones.
- Explore existing medical and synchrotron LDCT datasets and select which one(s) will be used. Consider using datasets that will allow comparing results with existing methods.
- Train diffusion models over synchrotron CT images and test over synchrotron LDCT ones.
- Train diffusion models over medical CT images and test over the synchrotron LDCT ones.
- Compare denoising performance with other existing methods.
- Evaluate and test possible ways to enhance acquired results.

This work aims to contribute with a methodology based on diffusion generative models to help enhancing synchrotron LDCT image quality towards denoising it. Its is part of its goal to explore the possibility of taking advantage of medical CT datasets to train models to be used over synchrotron LDCT images as they are more abundant than synchrotron CT datasets.

1.5 Text Organization

This project is structured into 5 chapters. Chapter 1 delineates the research problem under investigation, along with articulating the principal objectives and expected contributions. Chapter 2 provides an exposition of pertinent concepts essential for grasping the dissertation's scope, accompanied by a review of relevant literature. Chapter 3 elucidates the methodology employed, alongside detailing the datasets utilized and the metrics adopted for evaluating the models. Chapter 4 outlines the work plan and execution timeline.

Chapter 2

Literature Review

In this chapter, the literature review conducted for this work is presented. Section 2.1 covers the main concepts and techniques related to this work and Section 2.2 provides an analysis of relevant developed works related to LDCT denoising.

2.1 Concepts and Techniques

This section covers a wide range of relevant concepts and techniques related to LDCT imaging and denoising, such as: Electromagnetic (EM) radiation attenuation, CT image acquisition, reconstruction and quality assessment, ML, Artificial Neural Network (ANN), CNN and different architectures of generative models.

2.1.1 Electromagnetic Radiation

Bushberg et al. [9] defined radiation as "energy that travels through space or matter". EM radiation is usually characterized by its wavelength(λ), frequency(ν) and energy per photon (E). Variations on those features creates the so called "EM spectrum" which is commonly categorized into groups (microwave, Infrared (IR), Ultraviolet (UV), X-Ray, etc.) for convenience. Due to the "wave-particle duality" concept from quantum mechanics, EM radiation can be described as both waves and particles called *photons* [9].

For CT imaging domain, the X-Ray spectrum is the used one, being also sub-divided in Soft X-Ray ($E \leq 10keV$) and Hard X-Ray ($E > 10keV$). Additionally, depending on photon's energy, its interaction may cause ionization on the interacted molecule/atom, in other words, it may cause it to lose or gain one or more electrons.

That said, EM radiation that causes ionization is called ionizing radiation. The ionization energy threshold varies among different molecules/atoms, meaning that the same energy may ionize a sample while may not ionize other [9]. Figure 2.1 illustrates an overview about the EM radiation spectrum.

2.1.2 Radiation Interaction with Matter

When EM radiation travels though matter, depending on photon energy and the material composition, it may penetrate without interaction or interact with it and be scattered, absorbed or transformed into particulate radiation on very high energy cases [9].

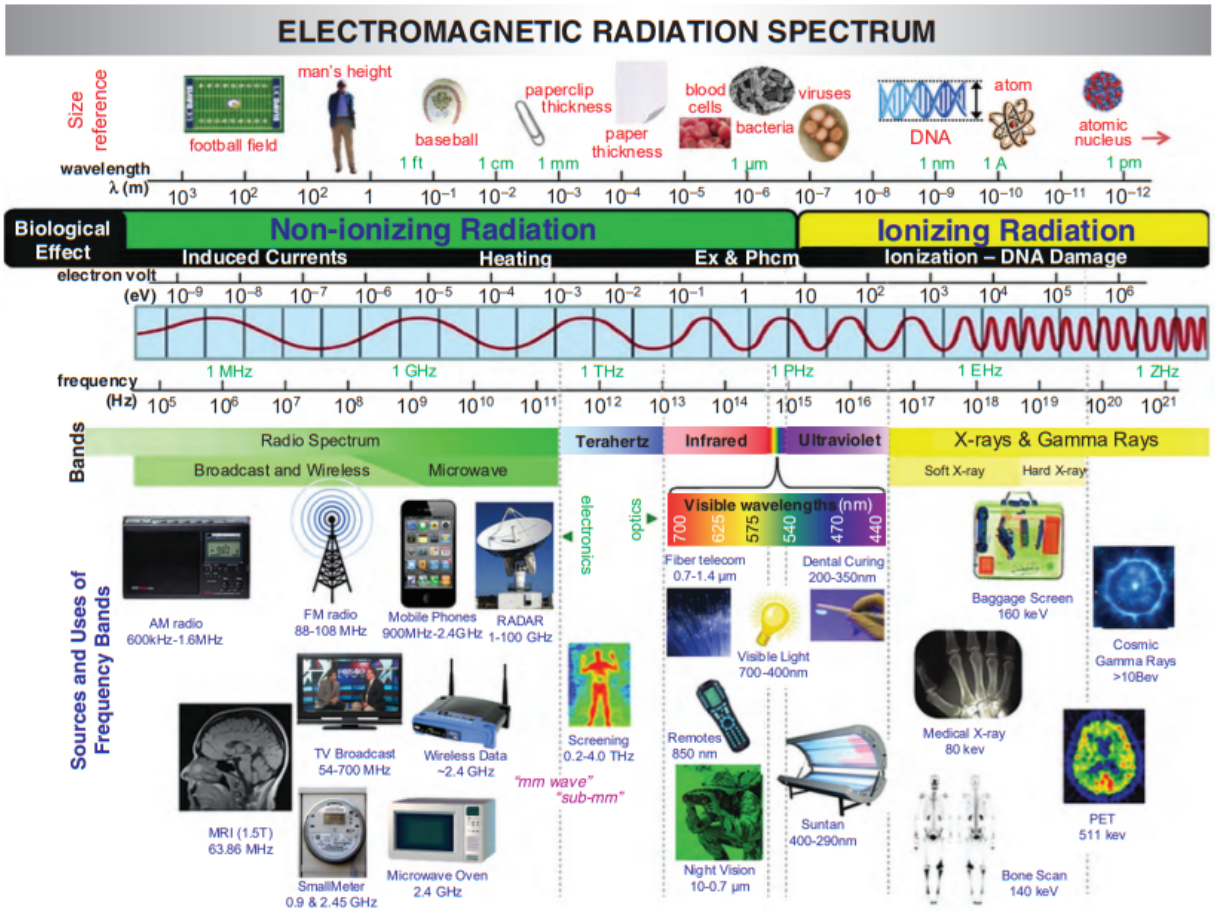


Figure 2.1: Overview of the EM radiation spectrum and its main characteristics. The ionization information varies depending on the irradiated object, in this example water is the referential molecule as most known living beings are mostly composed by water [9].

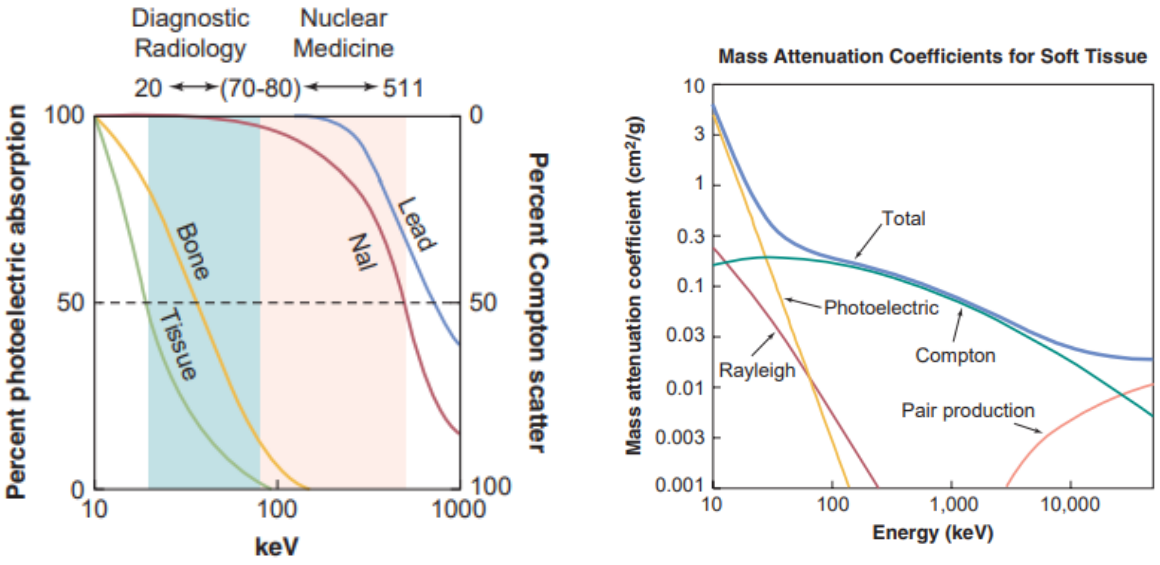
Scattering occurs when photons interacts with the whole atom (Rayleigh Scattering) or with electrons from atom's valence shells (Compton Scattering) deflecting their original trajectory, while absorption is when the photon energy is transferred to the atom and cause vibrations, excitation or even the ejection of electrons by what is called as photoelectric effect. Another interaction is called Pair Production, and it occurs during the interaction of photons with the electric field of a atom nucleus and results in the conversion of its energy into electron-positron particle pair [9].

That said, attenuation on X-Ray is caused by the combined occurrence of the previously explained interaction when radiation penetrates matter. Variation on materials composition and thickness changes the interaction probability, making it possible to define it as a proportion of atoms per volume, which is called *Mass Attenuation Coefficient* defined in Equation 2.1, where μ is the linear attenuation coefficient of the material and ρ its density. Explaining a little more about μ, it can be defined as the sum of individual linear attenuation coefficients for each type of interactions, as shown in Equation 2.2. As highlight previously, the occurrence of each type o interaction also varies depending on photon energy, which means that, different materials have different attenuation profile curves based on EM radiation energy as illustrated in Figure 2.2 [9]. Those attenuation curve profiles

can be faced as unique signatures of each material.

$$\text{Mass Attenuation Coefficient} = \frac{\mu}{\rho} \tag{2.1}$$

$$\mu = \mu_{\text{Rayleigh scatter}} + \mu_{\text{photoelectric effect}} + \mu_{\text{Compton scatter}} + \mu_{\text{pair production}} \tag{2.2}$$



(a) (b)

Figure 2.2: (a) Variation in the occurrence between photoelectric absorption and Compton scatter for different materials and energies. (b) Variation of the mass attenuation coefficient for soft tissues for different energy values [9].

2.1.3 Synchrotron Light Sources

Synchrotron light is produced when charged particles, like electrons, are accelerated to nearly the speed of light and compelled to move in circular or helical trajectories by powerful magnetic fields. As these particles alter their direction, they emit EM radiation tangentially to their path, called synchrotron light, which covers a wide EM spectrum, ranging from infrared to X-Rays [55].

Due to that phenomena, synchrotron facilities are an extreme versatile tool for science as it allows to perform a broad range of experiments over a big diversity of samples resulting in its usage by researchers from many different disciplines as illustrated in Figure 2.3. It is also important to highlight that some types of experiment are also only possible due to synchrotrons light sources as the huge amount of photons generated by it highly surpass standard laboratory X-Ray sources making experiments to be significantly faster [48].

Synchrotron facilities are typically classified by generations based on their brilliance. Currently, the most advanced are 4th generation facilities, which include Sirius in Brazil, European Synchrotron Radiation Source (ESRF) in France, and MAX-IV in Sweden. Other countries are also upgrading or building 4th generation synchrotrons. For example, Advanced Photon Source (APS) in the United States of America (USA) is being upgraded to become a 4th generation facility, and High Energy Photon Source (HEPS) in China is under construction.

Figure 2.3: Synchrotron users by discipline. Figure from Willmott [55] which estimated the data from user data between 2013 and 2016 from 4 synchrotrons facilities: SLS, SSRL, APS ans ESRF.

2.1.4 Radiation Dose

Initially, it is important to briefly define the concepts of *fluence* (ϕ) and *flux* (ψ). Fluence is the amount of photons/particles passing through an area as defined in Equation 2.3 while flux, for a mono-energetic beam of photons, is the product of ϕ and the energy per photon E, which can be interpreted as the amount of energy passing by an area, as defined in Equation 2.4 [9].

$$\phi = \frac{Photons}{Area} \tag{2.3}$$

$$\psi = \phi \times E \tag{2.4}$$

Radiation dose is related to the energy deposited in the irradiated material by the many types of interactions explained previously in Subsection 2.1.2. The Absorbed Dose (D) can be defined in terms of the energy (E) and the mass of the irradiated material (m), as shown in Equation 2.5 [9].

$$D = \frac{E}{m} \tag{2.5}$$

However, it is important to highlight that Equation 2.5 gives a theoretical and simplified way to calculate the dose over a material. In reality, if we consider the poly-energetic behavior of radiation sources, heterogeneity of materials, how different types of radiation interacts with matter and also the possibility of energy that was scattered being scattered back when interacting with nearby materials and interacting again with the same material, the calculation of dose gets much more complex and really specific to the radiation type and energy, material morphology and composition, and the environment/surroundings where the exposition occurs [9].

Due to that complexity, the calculation of dose differs depending on the application. Examples are the definition of Equivalent Dose (H) and Effective Dose (E_{dose}), which are commonly used for radiation protections and medical imaging applications. On Equivalent Dose, it is defined an a weight factor (w_R) in order to model the effectiveness of the type of radiation in causing damage to the material, as defined in Equation2.6. In the other hand, Effective Dose (E_{dose}) tries to model the sensitivity of materials and give a specific weight for each of them, this way, it is defined as the sum of each Equivalent Dose (H_T) on each

different material weighted by its sensitivity (w_T) as defined in Equation 2.7 [9].

$$H = D \times w_R \tag{2.6}$$

$$E_{dose} = \sum_T (w_T \times H_T) \tag{2.7}$$

Finally, it is important to highlight that depending on the dose a living being is exposed to, it may cause severe health problems and even its death [9, 40] and on non-living but radiation-sensitive samples it may lead to its degradation and directly affect the acquired data [36].

2.1.5 X-Ray Computed Tomography (CT)

Medical and synchrotron CT imaging, in a simplified way, can be explained by two main steps: data acquisition and reconstruction. The data acquisition step involves irradiating the object, whose 3D morphology will be analyzed, with X-Rays from many different angles. This can be done by rotating the X-Ray source around a stationary sample (more common in medical imaging), or by rotating the sample while the X-Ray source remains fixed (more common in synchrotron CT). For each angle to which the sample is exposed, due to the attenuation caused by the interaction between the EM radiation and the sample, a detector aligned with the X-Ray source positioned behind the sample collects transmission projection images from it. With many projections acquired from the sample from different angles, the challenge is to reconstruct the 3D morphology of the object using these projections.

The reconstruction can be formulated as an inverse problem, considering $\mu(x, y)$ as the original, but unknown, attenuation coefficient distribution of the object, being x and y spatial positions. The challenge is to discover $\mu(x, y)$ given the information acquired from N projection images $p_\theta(t)$ of the object, where θ is the respective angle at which it was taken and t is the position on the detector where the projection attenuation measurement is recorded.

The acquired projections can be considered as line integrals of $\mu(x, y)$, meaning that for a given angle θ, the projection $p_\theta(t)$ is given by the Radon transform presented in Equation 2.8, where δ is the Dirac delta function.

$$p_\theta(t) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mu(x, y)\delta(t - x\cos\theta - y\sin\theta)\, dx\, dy \tag{2.8}$$

The problem is defined as inverse because the goal can be interpreted as calculating the inverse of the Radon transform. To solve this, analytical algorithms such as Filtered Back Projection (FBP) or iterative algorithms are commonly used to approximate the Radon inverse transform and allow reconstructing $\mu(x, y)$, which further allows acquiring slice images from the object that can later be analyzed or used to create a single 3D volume of the object.

In theory, using an infinite number of projections with a minimum angle difference between each would allow for the exact reconstruction of the material. However, this is impractical due to physical, time, and dose limitations. Generally, acquiring more projections results in a better-reconstructed image with less noise. Conversely, a lower number of projections leads to more degraded images with higher noise levels, as there is less signal information available.

2.1.6 Low-Dose Computed Tomography (LDCT)

The development of LDCT techniques is crucial in the medical imaging domain to reduce health risks caused by radiation exposure. In synchrotron experiments, LDCT concerns typically involve the analysis of in-vivo samples or radiation-sensitive samples. High doses can damage these samples, affecting measurements or even precluding the overall experiment [36].

There is a clear trade-off between dose reduction and image quality, as reducing the dose generally leads to lower image quality. Therefore, LDCT acquisition protocols usually follow the ALARA principle, which aims to expose the patient or object under study to the minimum radiation necessary to achieve results of sufficient quality for the intended analysis.

Defining a precise dose threshold is extremely complex because dose absorption varies greatly based on material composition, morphology, and EM energy, as detailed in Subsections 2.1.2 and 2.1.4. Consequently, LDCT imaging protocols are often variations of Normal-Dose Computed Tomography (NDCT) protocols but with some dose mitigation strategies.

Finally, it is important to ensure that the dose reduction in LDCT techniques does not degrade image quality to the extent that it precludes the intended analysis. Such degradation would necessitate a repeat exposure, increasing the total dose and violating the ALARA principle.

2.1.7 Digital Image Representation

An image can be understood as an 2D signal and be simply defined by a function f where for a given set of spatial coordinates (x, y) a given intensity is mapped [20]. That said, continuous image representation is mathematically modeled as a function $f(x, y)$ where (x, y) are the spatial coordinates from a continuous spatial domain, and f assigns intensity values from a continuous range to each point that was acquired by an imaging system. In order to convert the continuous image $f(x, y)$ to its digital version two processes called sampling and quantization must be performed. Sampling is basically the process to express the continuous spatial coordinates (x, y) into discrete coordinates (i, j) by mapping Δ_x and Δ_y intervals into respective (i, j) pairs, which causes the effect of partitioning the xy-plane into a grid [20]. Quantization is the process to transform the continuous intensities given by $f(x, y)$ into a discrete set of values fitting the discrete grid defined by (i, j) [20]. After doing both sampling and quantization, a new $f(i, j)$ discrete image is created which is basically the digital version of the $f(x, y)$ continuous image. For 3D images, which may also be called as volumes, this whole process is the same but must be extended to cover the third dimension.

2.1.8 CT Image Quality

CT image quality is strongly bounded to the definition of three types of resolution: Spatial (or High-Contrast), Contrast (or Low-Contrast) and Temporal Resolution [9].

Spatial resolution is the ability to differentiate physically near objects in the CT image. It may also be called as High-Contrast Resolution as this differentiation is related to objects with different densities, which in general, results in higher grayscale contrast between

objects. Therefore, it focus on how well-defined and sharp are edges and details from the objects [9, 20].

Contrast resolution do not focus on objects spatial proximity but on being capable to discern similar density objects based on their grayscale values. Due to the density similarity, the grayscale values for those objects are usually very similar, making contrast resolution also be called as Low-Contrast Resolution, as it focus on how differentiable are object of similar grayscale values [9, 20].

Temporal resolution refers to the time required to acquire data, specifically how quickly CT image acquisition can occur. Low temporal resolution can impact image quality, particularly in cases involving moving subjects such as imaging living organisms, as movements like breathing or heartbeats can cause position variations, leading to fluctuations in attenuation over time [9].

Both spatial and contrast resolutions are strongly affected by noise, as it directly degrade the grayscale intensities and worse edges and details within the image. Given that, working on strategies able to mitigate or treat noise over the whole CT imaging pipeline is an extremely important effort to enhance CT image quality.

2.1.9 Machine Learning

ML is a subset of Artificial Intelligence (AI) that focuses on the development of algorithms and statistical models that enable computers to perform specific tasks effectively. The main difference of ML among other AI system are that those tasks are accomplished by learning from patterns and inferences derived from data, rather than following explicit instructions written by humans.

2.1.10 Artificial Neural Networks

ANN is a research area within ML which originally was inspired by studies over how information is processed and propagated on mammals brains [5]. Many mathematical models were proposed but the formulation expressed in Equations 2.9 and 2.10 converged to be the foundational form for ANN models until present days [5]. For a better understanding, Equation 2.9 model the interaction of a neuron with other M neurons as a linear combination where x_i represents the signal propagated by each other neuron and w_i the weight of those signal defined by the synapse connection, which results in a value a called *pre-activation*. After that, the *activation* of the neuron is defined in Equation 2.10 where a non-linear function f called *activation function* is applied over a resulting in what is called the neuron *activation* represented by y. If this neuron is connected to other, y would be the x value propagated by that neuron for a next neuron it is connected to. A common diagram used to present the behavior expressed in Equations 2.9 and 2.10 is shown in Figure 2.4 [5].

Creating layers with multiple neurons together, stacking them into multi-layers architectures and varying the *activation function* used, gives the capability of the model to represent very complex linear and non-linear functions, making ANN a flexible and multi-purpose approach to solve a big diversity of problems.

$$a = \sum_{i=1}^{M} w_i xi \tag{2.9}$$

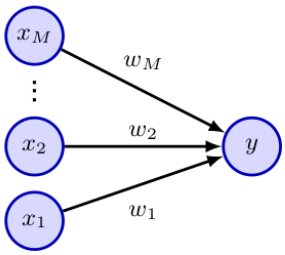


Figure 2.4: Graphical representation of ANN neurons [5].

$$y = f(a) \tag{2.10}$$

Bishop and Bishop [5] categorized the advances in ANNs into three phases. The first phase features single-layer networks, epitomized by the perceptron [41]. The second phase involves multi-layer networks, enabled by error backpropagation strategies and gradient-based optimizations, but hampered by ineffective and computational costly training of networks with more than 2 layers, which required specialized feature extraction pre-processing steps to be applied. The third and current phase, emerged in the last decade and encompasses Deep Neural Networks (DNNs) with models with up to one trillion of trainable parameters, contrasting drastically with the hundred to thousands parameters from previous phase [5].

DNN models became able to work directly over raw data as their numerous hidden layers allow the models to automatically extract hierarchical features from raw data, progressively capturing higher-level abstractions with each layer. DNN success was directly driven by advancements in computational power, availability of large-scale annotated datasets and more sophisticated training algorithms. Over many years different DNN architectures were develop to deal with specific types of data and challenges creating specialized classes of DNN to enhance their results over those specific domains [5].

2.1.11 Convolutional Neural Networks (CNNs)

CNNs [3,5,33] represent a specialized class of neural networks tailored for processing image data. Unlike traditional neural networks, which require a vast number of parameters to handle image data, CNNs exploit spatial hierarchies using convolutional layers. This approach significantly reduces the number of parameters needed and lowers the computational cost during training.

Convolutional layers within CNNs are responsible for learning image filters, also known as kernels, which detect local patterns such as edges, textures, and shapes relevant to the model's target task. These kernels are applied over the input image through convolution operations, generating feature maps. Subsequently, pooling layers downsample these feature maps, summarizing them into single values through fixed and not learned operations such as max, min, or average pooling.

The ability of CNNs to handle tasks directly from raw image data has revolutionized solutions in CV tasks and has also been extensively explored in challenges involving other data dimensions such as temporal series, volumetric images, and videos [5].

2.1.12 Latent Space Variables and Data Space Variables

Latent Space Variables $\mathbf{z}$ are values that are assumed to exist in a distribution but were never observed, while the data/observed space variables $\mathbf{x}$, are values really acquired or, in other term, observed [4, 5]. Latent space variables are important because they are a foundational concept of latent-variable models which are also a core concept of most generative models.

Latent Space Variable are used in order to allow models to learn transformation functions capable of transforming a variable $\mathbf{z}$ from a latent space distribution $p(\mathbf{z})$ to a variable within the data space distribution $p(\mathbf{x})$, allowing those model to generate new samples that would follow the Probability Density Function (PDF) $p(\mathbf{x})$ [5].

2.1.13 Linear and Non-Linear Latent-Variable Models

Linear latent-variable models are statistical tools used to uncover hidden (latent) structures within observed data. These models assume that the observed variables $\mathbf{x}$ are influenced by a smaller number of unobserved, or latent, variables $\mathbf{z}$, which in turn follow a linear relationship which, if known, can be used to transform variables from latent space $p(\mathbf{z})$ into observed variable from data space $p(\mathbf{x})$, and vice-versa [4, 5].

Even though, modeling variables as linear combinations of latent variables may help to simplify complex data and ease to analyze and interpret underlying patterns, the assumption of linearity is too restrictive and cannot adequately capture the complexity of the data on many cases. Due to that, non-linear latent-variable models extend the concept previously defined by linear latent-variable models and consider that a non-linear relationship is followed [4, 5].

The covariance structure of the observed variables in the context of a non-linear latent-variable model is more complex and generally does not have a closed-form expression like in linear models. Due to that, the usage of ML approaches and more specifically ANN has been vastly explored to learn a approximations of the non-linear relationship between $p(\mathbf{x})$ and $p(\mathbf{z})$ variables [5].

The concept of learning a way to represent latent variable z from a latent space $p(z)$ to a variable x on data space $p(x)$ is a core concept which generative models are built over as illustrated in Figure 2.5 which gives high level overview of different generative models.

2.1.14 Generative Models and Discriminative Models

ML models are commonly classified as generative or discriminative. Generative models are a group of ML models that focus on learning the PDF from a given train set in order to be able to generate new examples from that distribution [5]. In contrast, discriminative models aim to learn a function able to define boundaries capable of discriminating data points from different classes within a data distribution. Figure 2.6 illustrates some differences among discriminative, generative, and conditioned generative models.

Putting this in mathematical terms, consider a vector of data points $\mathbf{x}$ in the data space and their respective classes $\mathbf{y}$. It is possible to define that each data point from $\mathbf{x}$ belongs to a given PDF $p(\mathbf{x})$ and, in the same way, each label from $\mathbf{y}$ to a PDF $p(\mathbf{y})$.

The goal of discriminative models is to learn the PDF of a label $\mathbf{y}$ to occur given a data point $\mathbf{x}$, which can be denoted as a PDF $p(\mathbf{y}|\mathbf{x})$. In the other hand, generative models goal is to learn how to replicate the $p(\mathbf{x})$ PDF so it will be able to synthesize data within that

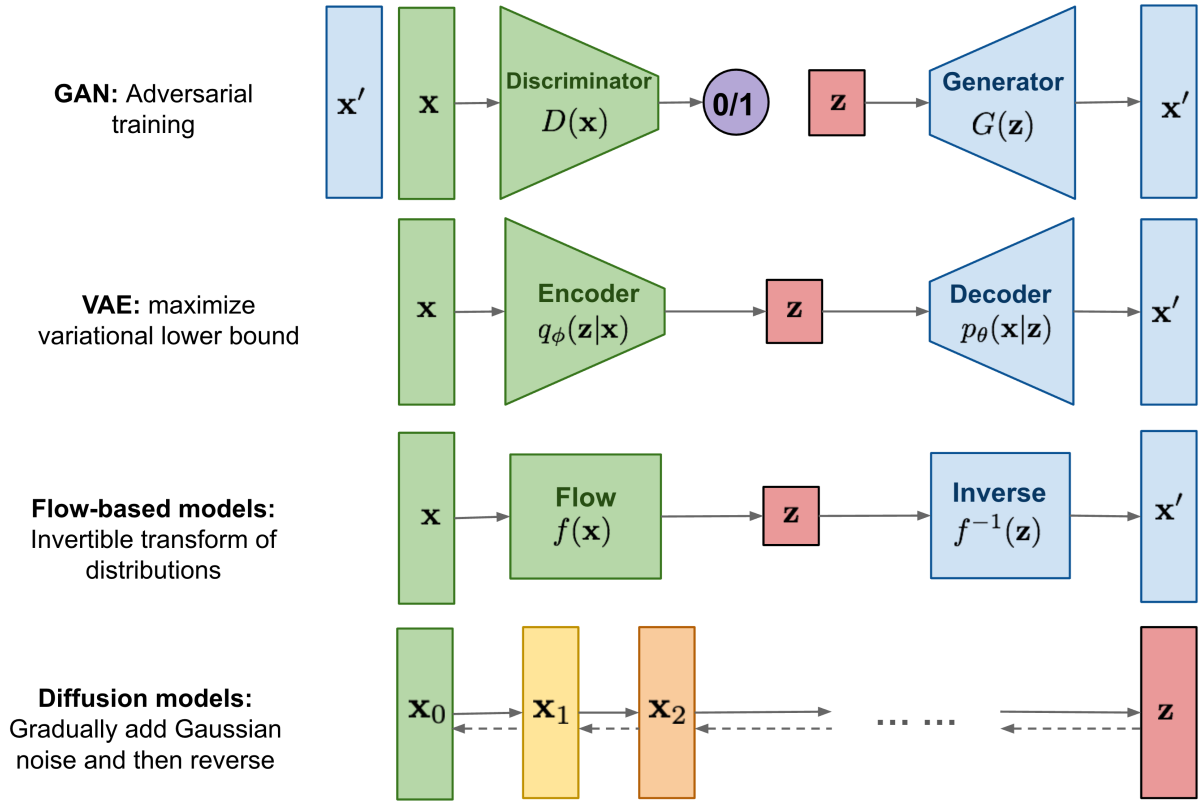


Figure 2.5: High level overview of different types of generative models. All of the 4 types uses the concept of generating new sample by transforming a latent variable $\mathbf{z}$ into a data space variable $\mathbf{x}'$ [54].

data distribution. For that, a latent space variable $\mathbf{z}$ from a latent distribution $p(\mathbf{z})$ is used, in reality, what generative model learns is the PDF $p(\mathbf{x}|\mathbf{z})$, which in other words can be defined as, how to transform a latent variable $\mathbf{z}$ to a data space variable $\mathbf{x}$. Ideally, if the model learns perfectly, $p(\mathbf{x}|\mathbf{z})$ would be exactly $p(\mathbf{x})$.

In many applications, it becomes useful the possibility to condition the generative model output based on a given class. In that case, the generative model aims to learn the PDF of a data point $\mathbf{x}$ for a given latent variable $\mathbf{z}$ and its class $\mathbf{y}$ which can be denoted as the $p(\mathbf{x}|\mathbf{z}, \mathbf{y})$.

Bishop and Bishop [5] defined that generative models can be denoted as a distribution, where it returns the probability of a data point $\mathbf{x}$ based on its learnable parameters $\mathbf{w}$ denoting it as $p(\mathbf{x}|\mathbf{w})$. Similarly, conditional generative models could be define as $p(\mathbf{x}|\mathbf{c}, \mathbf{w})$ where $\mathbf{c}$ is a vector that defines the conditioning variables to guide the generation of $\mathbf{x}$.

Additionally, according to Bishop and Bishop [5], in principle, ANN based generative models should be capable of approximating any desired distribution, due to the universality capability of ANN transformation.

2.1.15 Generative Adversarial Networks (GANs)

GANs were firstly proposed by Goodfellow et al. [21] and represented an innovative approach for generative models due to the usage of two competing ANN on its architecture: the generator $g(\mathbf{z}, \mathbf{w})$ and the discriminator $d(\mathbf{x}, \boldsymbol{\phi})$.

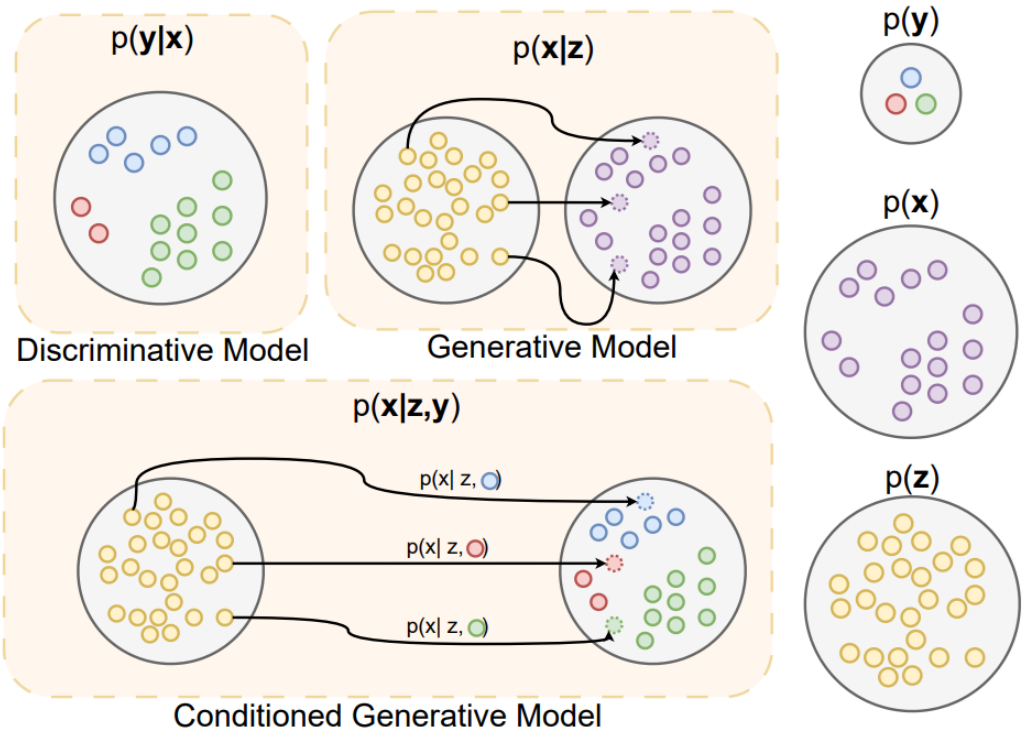


Figure 2.6: Differences among discriminative, generative, and conditioned generative models.

The generator aims to create realistic data samples by learning how to transform variables from a latent space $p(\mathbf{z})$ into variables from the data space distribution $p(\mathbf{x})$. In contrast, the discriminator evaluates the generated samples against real samples, striving to accurately distinguish if the sample is fake or real. Figure 2.7 presents a visual example of GAN with both its generator and discriminator models.

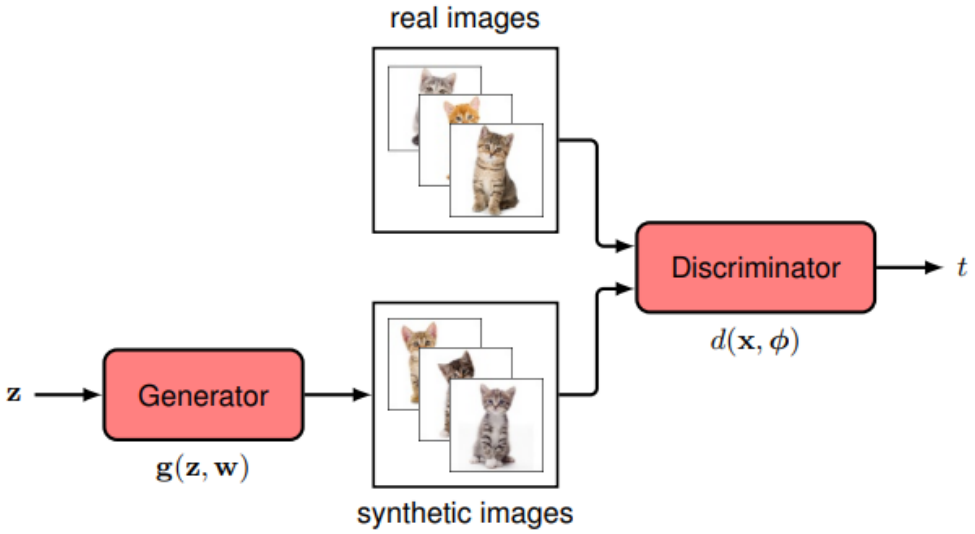


Figure 2.7: Visual example of GAN model composed by its generator and discriminator models [5].

During training, both models share the loss function, whose metric used on standard GANs architectures is usually binary cross-entropy. The error value is related to how good the discriminator correctly classifies the data samples on fake or real, the difference is that while the discriminator tries to minimize this error the generator aims to maximize it. That minmax dynamic results in both models playing what is know on game theory as a *zero-sum game* where the gain of one represents a loss to the other. This adversarial process drives both models to improve iteratively: the generator becomes adept at producing increasingly convincing fake data, while the discriminator sharpens its ability to detect forgeries [5].

One of the primary challenges in training GANs is known as *mode collapse*. This occurs when the generator fails to learn how to map the distribution $p(\mathbf{z})$ to the entire distribution $p(\mathbf{x})$, instead learning the mapping to only a subset of it. Consequently, the generator produces only a limited variety of outputs, rather than the full range defined by $p(\mathbf{x})$.

Furthermore, training GANs often encounters instability due to the absence of a clear progress metric and the adversarial nature of training. That cause errors in GAN training to fluctuate unpredictably, complicating the process of achieving consistent improvement.

For those reasons, changes on the loss function used or the training procedure has been proposed in order to improve training stability and also to try to guarantee that the PDF being learned by the generator is the wanted PDF.

One popular approach is the Wasserstein Generative Adversarial Network (WGAN) proposed by Arjovsky, Chintala and Bottou [37] which instead of using the standard binary cross-entropy to define model errors, applies the Wasserstein distance metric as the loss function which basically calculates how far apart are two distributions. Besides providing more stability during training, the usage of Wasserstein distance also ensure the generator learning process moves towards the desired distribution, also providing a meaningful loss metric able to show training progress and correlate with the quality of generated samples [5].

2.1.16 Deterministic and Variational Autoenconders (VAEs)

Autoencoders, also named as *auto-associative neural network*, are a known group of ANN models focused on learning internal representations of data. They are composed by two part, the first called *encoder* responsible for learning how to transform the input data into a new condensed representation and the second part, called *decoder*, focused on learning how to bring back the input data to its original representation [5].

Autoencoder models are usually classified into Deterministic or Variational and this is directly related to which is the new representation of data. For Deterministic autoencoders, a given data space variable $\mathbf{x}$ is transformed by a deterministic function $f(\mathbf{x})$ converting it into a new latent space variable $\mathbf{z}$, making the same $\mathbf{x}$ always be mapped to the same $\mathbf{z}$. In contrast, Variational Autoencoders (VAEs) learns a approximated posterior distribution $p(\mathbf{z}|\mathbf{x})$, commonly learned by a neural networks based *encoder* $e(\mathbf{z}|\mathbf{x}, \mathbf{w})$ where $\mathbf{w}$ stands for the model learnable parameters, and the *decoder* learns how to reconstruct $\mathbf{x}$ by learning the PDF $p(\mathbf{x}|\mathbf{z}, \boldsymbol{\phi})$ where $\boldsymbol{\phi}$ are the models learnable parameters [5].

Autoencoder models training is done by feeding the model with samples and evaluating if it is capable of generating it back. Obviously, reconstructing the exact input data is not possible but something very similar is desirable, so the train error is commonly related to the difference between the input data $\mathbf{x}$ and the output data $\hat{\mathbf{x}}$ [5].

2.1.17 Diffusion Models

Diffusion models [46] are a novel class of generative models that employ a process inspired by physical diffusion phenomena to generate data. These models typically involve two main phases: a forward encoder and a reverse decoder. In the forward encoder, data is gradually corrupted by the addition of Gaussian noise over several steps, effectively transforming the data into a noise-like distribution. The reverse process, often parameterized by a neural network, aims to progressively denoise the data, reconstructing it step-by-step from the noise. Diffusion models works very similar to VAEs but instead of training both the encoder and decoder, the encoding distribution is fixed and determined by the noise, while the decoding is learned during training and is done gradually [5].

Giving a more formal definition for the whole process. The forward encoder in diffusion models progressively corrupts a data sample $\mathbf{x}$ from a given data space $p(\mathbf{x})$ through a sequence of T steps. At each step t, noise is added to the sample, resulting in a more noisy version $\mathbf{z}_t$ of the previous $\mathbf{z}_{t-1}$ version.

The noise added at each step is defined in Equation 2.11, where β_t is the variance of the noise distribution and $\boldsymbol{\epsilon}_t \sim N(\boldsymbol{\epsilon}_t|0, \mathbf{I})$ being $\mathbf{I}$ the unit matrix. The variance of each step β_t receives values within the interval $(0, 1)$ and increase at each step following a prescribed schedule where $\beta_1 < \beta_2 < \cdots < \beta_T$. The usage of both coefficients $\sqrt{1-\beta_t}$ and $\sqrt{\beta_t}$ ensures that, for each step t, the mean and variance of $\mathbf{z}_t$ are, respectively, closer to 0 and to the unit matrix if compared to $\mathbf{z}_{t-1}$. Thus, considering $T \to \infty$, consequently $\beta_T \to 1$ and $\mathbf{z}_T$ would converge to be a variable within the prior defined Gaussian distribution of the noise $N(0, \mathbf{I})$, effectively causing the complete loss of all information from the data space and its conversion to a latent variable $\mathbf{z}_T$ [5].

$$\mathbf{z}_t = \sqrt{1-\beta_t}\mathbf{z}_{t-1} + \sqrt{\beta_t}\boldsymbol{\epsilon}_t \tag{2.11}$$

The decoding process can be defined as learning a Gaussian conditional distribution $q(\mathbf{z}_{t-1}|\mathbf{z}_t)$ which would be able to revert a latent variable $\mathbf{z}_t$ into its previous state $\mathbf{z}_{t-1}$, which could them be applied over and over again until restoring the original data space variable $\mathbf{x}$. Theoretically, a procedure similar to the forward encoder could be applied, however, on the forward encoder procedure, the target distribution $p(\mathbf{z})$ was known as it was directly defined by the noise applied, in the reverse decoding procedure, that is not possible as the data space distribution $p(\mathbf{x})$ is unknown.

Due to that a DNN is trained to condition a reverse distribution $p(\mathbf{z}_{t-1}|\mathbf{z}_t, \mathbf{w})$ in order to approximate it by optimizing a likelihood metrics, where $\mathbf{w}$ are the learned parameters of the network. Intuitively, in order to make predicting the distribution $q(\mathbf{z}_{t-1}|\mathbf{z}_t)$ easier, if the variance of the noise add between steps is really small ($\beta_t \ll 1$), the inversion should be easier to learn as difference between distributions $q(\mathbf{z}_{t-1})$ and $q(\mathbf{z})$ would be narrower. The downside of this is that T would get larger, making the process of generating new samples more costly, as all denoising steps must be done to generate a sample from the data space distribution $p(\mathbf{x})$ from $\mathbf{z}_T$. In contrast, the lower the value of T, more error is add on each step t causing the generated output to shift from the original sample $\mathbf{x}$ [5]. Figure 2.8 illustrates this gradual encoding and decoding procedure done by diffusion models.

One diffusion model variation reported by Bishop and Bishop [5] to bring higher quality results and firstly proposed by Ho et al. [24], is to make the ANN learn to predict the total noise added to the original sample in a given step t instead of predicting the denoised

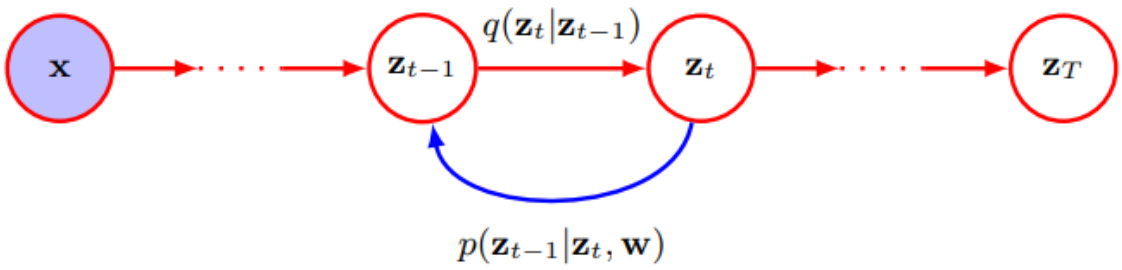


Figure 2.8: Illustration of the forward encoder and reverse decoder process of diffusion models [5].

version of the sample on each step.

Another interesting variation is Cold Diffusion, proposed by Bansal et al. [2], which discuss about the possibilities of not being a requirement for the forward process the addition of noise but instead that it is possible to use other types of arbitrary transformation.

Kingma et al. [30] showed that diffusion models can be view as stochastic differential equation SDEs, creating the theoretical basis that drives many approaches for achieving faster sampling speeds.

2.1.18 Low-Dose Computed Tomography (LDCT) Denoising

Methods to improve LDCT noisy images can be split in three groups [12, 36]. The first is related to approaches that tries to remove noise directly from the raw acquired data. Those approaches can be directly related to the acquisition process, calibration or also applying denoising algorithms over sinogram data. The second group is related to approaches that aims to improve reconstruction algorithms, focusing on being able to generate more precise and less noisy reconstructed images. The last, and most commonly explored, are methods which tries to perform denoise as a post-processing over already reconstructed images.

The main goal of LDCT denoising tasks is to develop a solution capable of removing the noise in the image and restoring the original signal values in it, but it is also important that this solution is capable of doing that while not degrading structural information from the image. Because of this, two common metrics used to evaluate the quality of LDCT denoising solutions are the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) that allow to evaluate how much the signal and the structure information are degraded after applying the proposed denoising approach.

2.2 Related Work

In this section, a brief review of works related to denoising LDCT images is presented. The reviewed approaches were classified in 4 classes: Non-Learning Classical methods, CNN-based methods, GAN-based methods and Diffusion-based methods. Non-Learning Classical methods class is related to approaches that uses classical computer vision approaches and don't rely on training some kind of ML model to tackle the denoising problem. CNN-based, GAN-based and Diffusion-based methods, as their names suggest, are related to solutions that focus on using a specific type of ML models to handle the task.

2.2.1 Non-Learning Classical Methods

Buades and Morel [8] proposed a denoising technique called Non-Local Means (NL-means) which, different from many methods that just consider near pixels in order to fix the noise corrupted value of a pixel, uses the entire image by calculating a weighted average of all other pixels. The weights are defined by how the pixel neighborhood of each pixel is similar to the neighborhood of the pixel being denoised, the more similar the higher the weight. Later on, Dabov et al. [13] proposed the BM3D, a algorithm that starts by grouping similar 2D patches from the image called *blocks* into 3D data arrays called *groups* and use them in a two-step procedure called Collaborative filtering, responsible to apply 3D transformations over the grouped *blocks* and return a estimation of their denoised values to their original positions. Figure 2.9 illustrates the complete processing pipeline proposed by the authors for the BM3D algorithm.

BM3D is considered the state-of-art of non-learning classical methods and is commonly used as a baseline comparison to ML based approaches due to the denoising quality and computational cost of its execution.

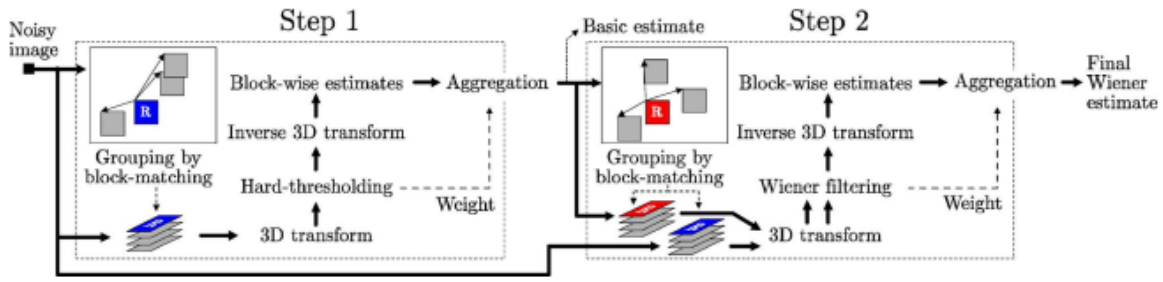


Figure 2.9: Processing pipeline applied by BM3D algorithm proposed by Dabov et al. [13].

2.2.2 CNN-based Methods

Chen et al. [12] proposed the RED-CNN (Residual Encoder-Decoder Convolutional Neural Network) model for handling the LDCT denoising task. It employs an encoder-decoder architecture with skip connections, which help to retain spatial information by passing feature maps directly from the encoder to the decoder. The residual learning approach used in RED-CNN further improves performance by allowing the network to focus on learning the residual noise, thus facilitating efficient and accurate noise reduction. The authors tested the model over clinical and simulated data against other non-learning classical models and other CNN-based approaches. An overview of the proposed RED-CNN architecture can be checked in Figure 2.10.

2.2.3 GAN-based Methods

One of the main reasons that guided GAN-based solution was seeking a solution to avoid oversmoothed edges and detail loss caused by CNN-based solutions.

Yang et al. [57] proposed tackling the LDCT denoising task by training a WGAN to learn how to map LDCT images denoised NDCT versions while adding to the loss function a perceptual similarity metric. The perceptual similarity metric is calculated over features from the generated and ground-truth images extracted by a pre-trained VGG network [45].

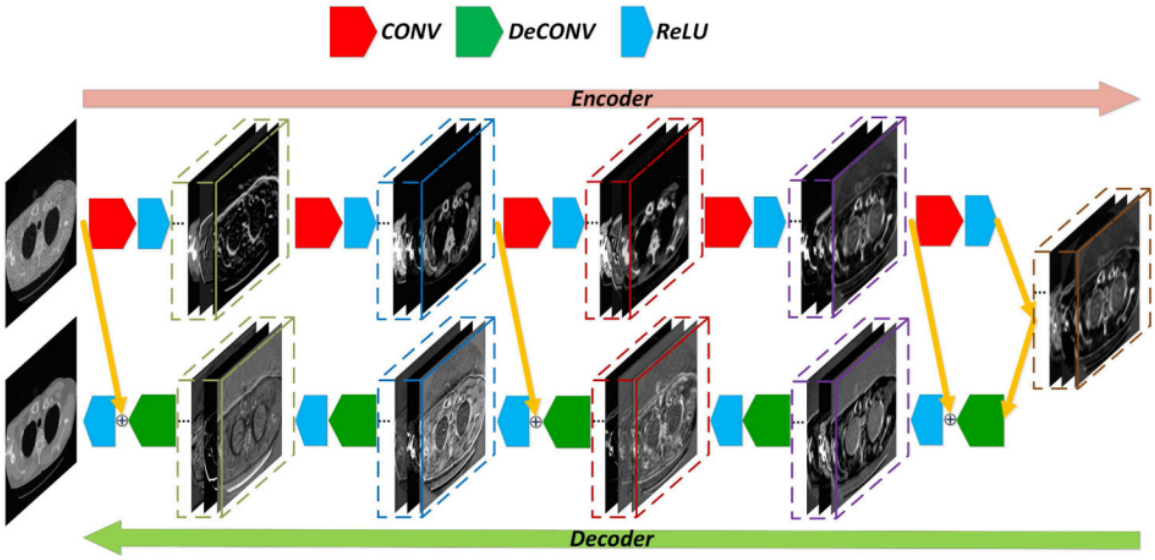


Figure 2.10: Overview of the RED-CNN model architecture proposed by Chen et al. [12].

Due to the usage of a WGAN and a VGG together, for convenience the model was named by the authors as WGAN-VGG network. It was reported that the WGAN-VGG was able to achieve less noisy and more detailed image when compared to CNN-based approaches.

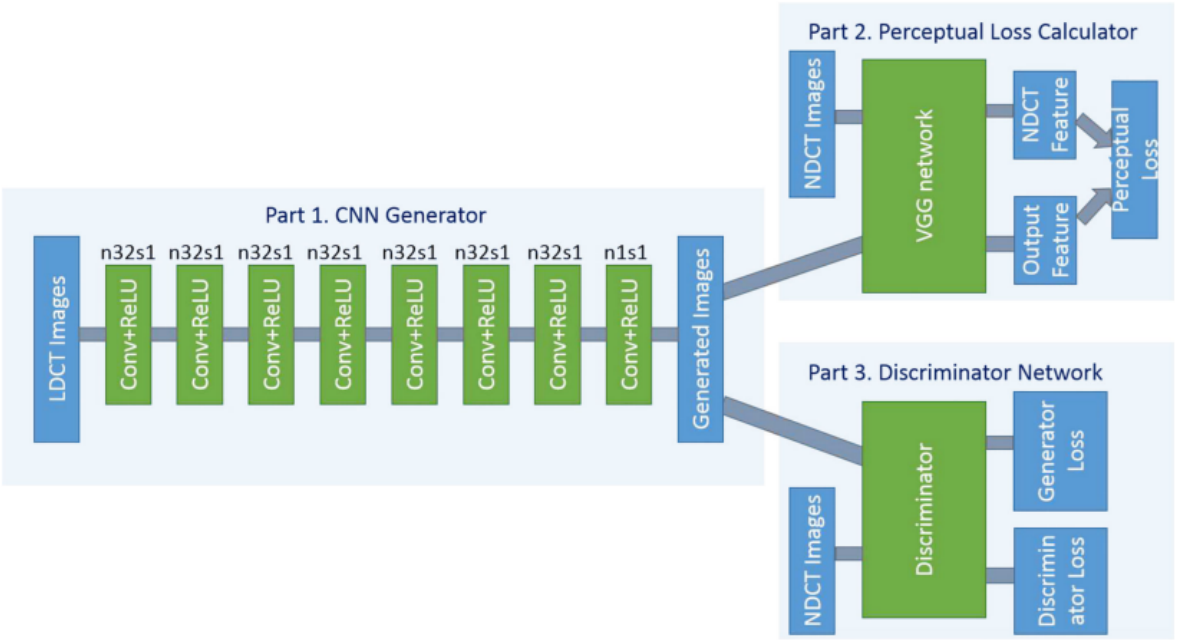


Figure 2.11: Overview of the WGAN-VGG model architecture proposed by Yang et al. [57].

Huang et al. [27] proposed DU-GAN which also aims to learn how to transform LDCT images into its NDCT version, but it uses two U-Net-based discriminators, that generates pixel-wise and global feedback about the image to the generator, which is a RED-CNN. One discriminator works in the image domain to force the generator to create realistic images while the second works in a gradient domain to encourage edge enhancement. Both discriminator errors are summed into a single one to be used into the models weight adjustment.

An overview of the DU-GAN model is illustrated in Figure 2.12.

The authors reported higher training costs but promising denoising and generalization results of DU-GAN against other GAN-based solutions with the additional feature of DU-GAN being able to output confidence maps.

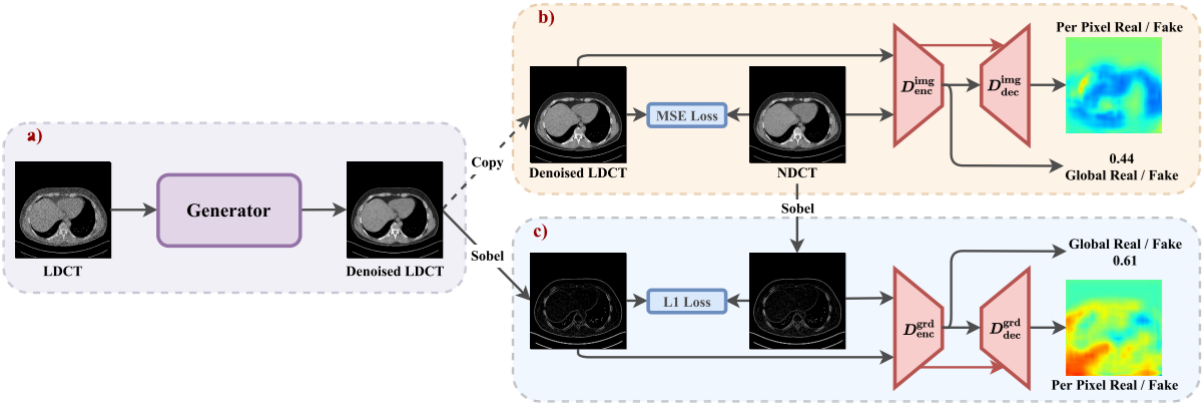


Figure 2.12: Overview of the DU-GAN model proposed by Huang et al. [27]. On (a) is presented the generator while on (b) and (c) the image and gradient domain discriminators, respectively.

2.2.4 Diffusion-based Methods

Xia et al. [56] proposed a solution for denoising LDCT images by training a DDPM model using NDCT and LDCT images. During training, the NDCT image is passed over the forward encoder which them is reconstruct by a reverse decoder conditioned by the LDCT version. The idea of using the LDCT version of the image to condition the denoising process is to guide the generation of a NDCT image related to the LDCT image. Using the proof developed by Song et al. [47], which shows that the reverse process of a DDPM has a equivalent form to a probability flow Ordinary Differential Equation (ODE), the authors also developed a fast ODE solver to speed-up the sampling process of their model, reporting to be 20× faster than the standard DDPM solution but still slower than a U-Net based solution. Nevertheless, their solution showed the best PSNR and SSIM scores among the compared solutions, showing to be better than a standard DDPM approach and also the U-Net approach in terms of denoising performance.

Gao and Shan [18] proposed a Contextual Conditonal Diffusion model (CoCoDiff) which focus on learning the noise add on each step t. Instead of giving the original image as the model input, a residual image is generated by subtracting the LDCT and NDCT images in order to ease model prediction. The residual image is used as the input for the forward encoder which corrupts it with Gaussian noise by t steps. During the reverse process the noise prediction model is used to define the noise to be removed on each step, which then is done interactively for each step t. In order to condition the noise prediction network to better respect 3D structural details, the NDCT image from current slice and both its adjacent slices are concatenated and passed as a condition to the noise predictor in order to deliver contextual information to it. Finally, a denoised version of the LDCT image can be generated by adding the residual image generate by the decoder in the LDCT image. A overview of the CoCoDiff model is presented in Figure 2.13.

The authors reported that CoCoDiff achieved better generalization over different dose levels when compared to the CNN and GAN-based approaches they tested against. Additionally, they acknowledged that the inference time of the model to be a limitation factor of the model, taking around 3.5 minutes for denoising a single 512x512 image using an NVIDIA V100 GPU.

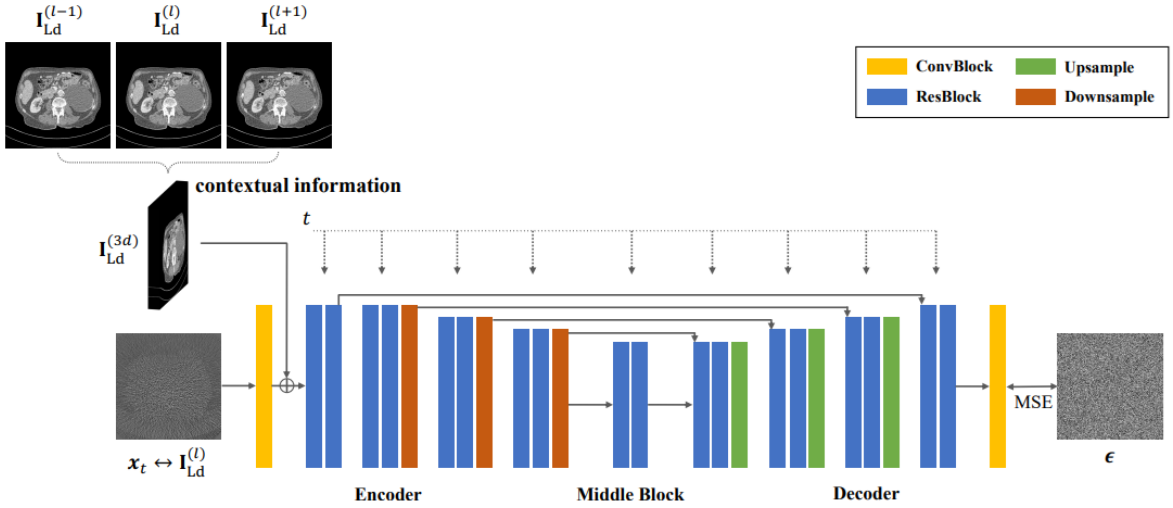


Figure 2.13: Overview of the CocoDiff model proposed by Gao and Shan [18].

Liu et al. [35] proposed an unsupervised approach that only uses NDCT samples during training to allow a zero-shot denoising method for LDCT images. The model developed by the authors is a two-step cascade diffusion model where the first model focus on generating lower resolution NDCT images from random noise and use them to condition a second diffusion model responsible for generating high-resolution NDCT images. This pre-trained two-step cascaded diffusion model is them used into a algorithm called Dn-Dp (Denoising with Diffusion Prior) proposed by the authors where the diffusion models priors are incorporated into a Maximum a Posteriori (MAP) framework to help solving each step of the LDCT image denoising. The authors compared their solution with other classical non-learning and learning based approaches, achieving the better denoising performance among unsupervised learning solution but losing against supervised learning approaches when noise were more pronounced. Nevertheless, the authors also report that their approach achieved visually more realistic results when compared to the oversmoothed results generated by supervised learning approaches and ineffectively of classical non-learning solutions.

Gao et al. [17] developed what they called as Contextual Error-modulated Generalized Diffusion Model (CoreDiff). Inspired by the Cold Diffusion model proposed by Bansal et al. [2], the authors proposed the usage of a new degradation operator during the diffusion process with the goal to mimic the physical degradation process that happens on CT images. Additionally, they developed a novel restoration network called Contextual Error-modulated Restoration Network (CLEAR-Net) in order to mitigate accumulated errors and misalignment caused by imperfect restorations networks. Finally, the authors also proposed a one-shot learning framework customized for CoreDiff to promote rapid model generalization for unseen dose levels. A overview of the CoreDiff model is illustrated in Figure 2.14.

A vast experimentation was done against non-learning classical algorithms, CNN-based, GAN-based and other diffusion-based solutions. CoreDiff was reported to show promising

denoising performance while requiring only 10 sampling steps making its inference time significantly faster than other classical diffusion approaches.

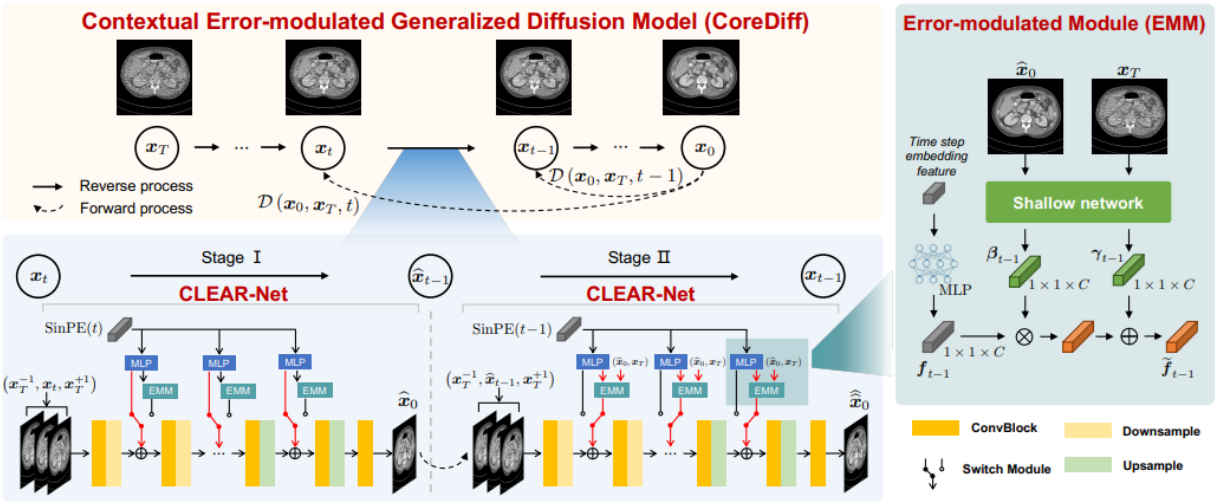


Figure 2.14: Overview of the CoreDiff model proposed by Gao et al. [17].

Notably, none of the diffusion-based approaches presented in this section were used over synchrotron CT images, reinforcing the innovative possibility of this work.

Chapter 3

Materials and Methods

In this chapter, the methodology proposed for this work is presented in Section 3.1 among with the datasets, metrics and computational resources that will be used in its development in Sections 3.2, 3.3 and 3.4, respectively.

3.1 Methodology

The proposed methodology for this work focus evaluating strategies based on applying diffusion generative models to perform denoising tasks over synchrotron LDCT images from varying dose levels. One of the strategies to be explore in the methodology is the evaluation of strategies to repurpose diffusion models trained over medical CT images to handle image denoising tasks over synchrotron LDCT images. The choice to try repurposing a model from another domain is the lack of public available synchrotron CT datasets and the abundance of medical CT ones. Additionally, due to the similarity between both domains, it is expected that strategies that works over medical LDCT images may work similarly well when applied to synchrotron LDCT images.

The capability of diffusion models to handle denoising tasks over medical images hasn't yet been vastly explored but had already achieved promising results. Among the diffusion-based solutions found during the performed literature review, CoreDiff [17] authors reported promising results in terms of image quality enhancement, inference time and model generalization among datasets and the availability of a one-shot learning mechanism for adapting the model for new unseen dose levels. Due to that, CoreDiff was chosen as the reference diffusion model for this work and the initial model architecture to be tested.

That said, the methodology was divided in 5 stages:

1. Dataset preparation;
2. Baseline metric calculation;
3. Experiment 1: Model trained over synchrotron CT images;
4. Experiment 2: Direct repurposed medical LDCT denoising model;
5. Experiment 3: Fine-tuned repurposed medical LDCT denoising model.

Figure 3.1 illustrates the procedure that will be performed on each stage which are also described in Sections 3.1.1, 3.1.2, 3.1.3, 3.1.4 and 3.1.5.

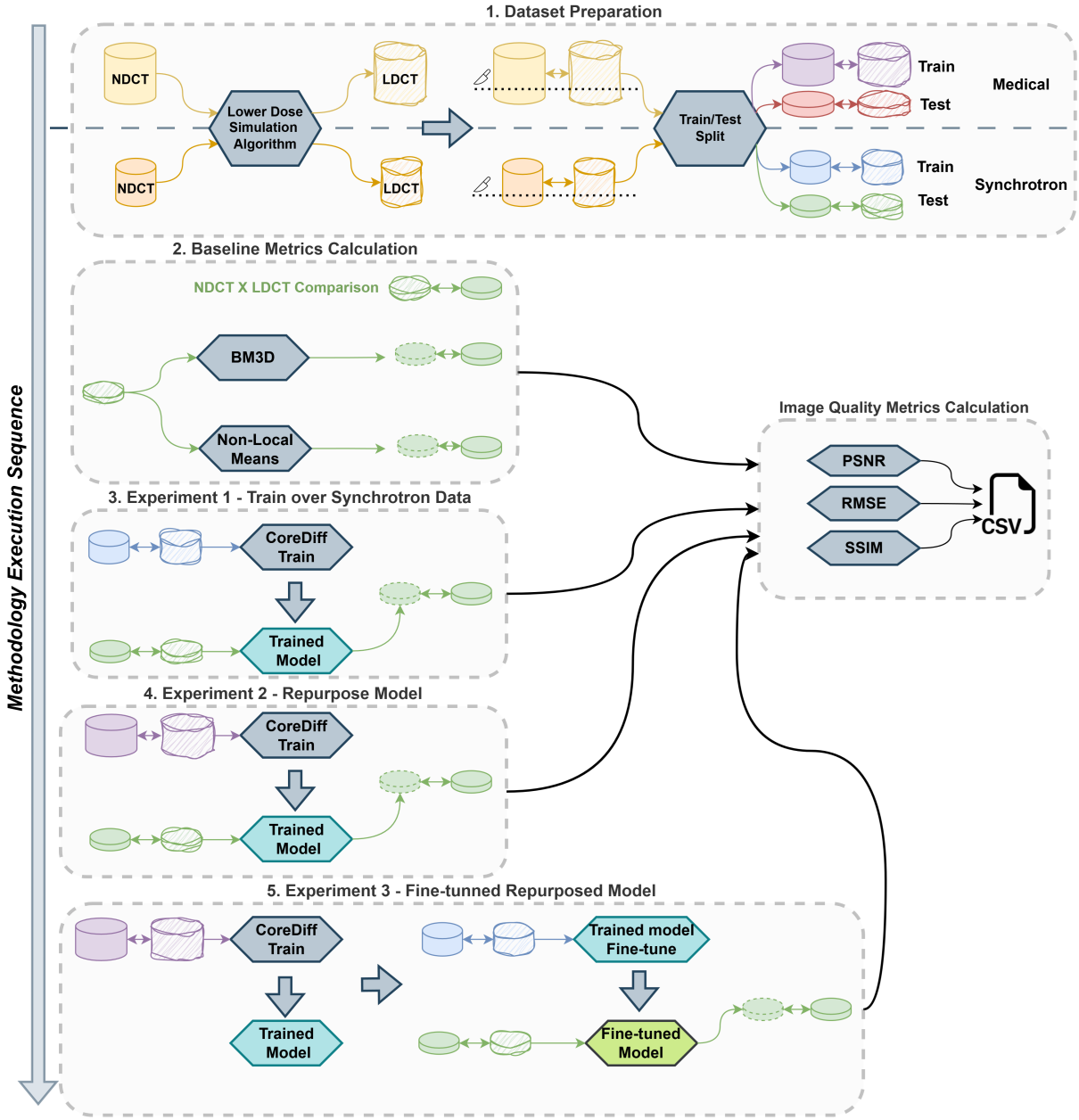


Figure 3.1: Proposed methodology. The stages are executed from stage 1 to 5 and the image quality metrics are calculated from the results acquired on stages 2, 3, 4 and 5 for further analysis.

3.1.1 Datasets Preparation

The available datasets include both medical and synchrotron NDCT images, with some also providing LDCT versions for specific reduced dose levels. In the medical CT domain, the object of study is in general the same, humans. Radiation dose applied may vary depending on body area being studied but in general, the object of study is known and allow CT scans calibrations. However, on synchrotron CT domain, the samples studied can range widely, going from minerals to biological samples, resulting in a diverse set of radiation dose constraints that differ from sample to sample.

Given these variations, it is important to test the model's performance across a broad range of dose levels for the synchrotron domain. This requires generating LDCT images at various dose levels for each NDCT image in the dataset. The dose variations in the lower dose images will be created using the simulation algorithm proposed by Yu et al. [58], which is widely used in medical CT datasets to generate LDCT images from NDCT scans. As a result, each sample in the dataset will include the original NDCT image along with LDCT images for each simulated dose level.

Finally, the entire dataset will be split into training and test sets to facilitate model training and performance evaluation.

3.1.2 Baseline Metric Calculation

To assess whether the proposed method improves the quality of synchrotron LDCT images, image quality metrics will be calculated by comparing the NDCT and LDCT versions of each image. These metrics serve as a fundamental benchmark; any method that produces worse results indicates that it degrades, rather than enhances, image quality.

Traditional non-learning techniques such as BM3D [13] and non-local means [8] will also be implemented and used as baseline comparison methods.

3.1.3 Experiment 1: Model Trained Over Synchrotron Data

In this experiment, the chosen diffusion model architecture will be trained using the train set from the synchrotron dataset and evaluated on its test set. The objective is to determine whether it is possible to train a model that can enhance the quality of synchrotron LDCT images despite the limited amount of available data. This will also enable a comparison between repurposed models and a model trained specifically for this domain with fewer data.

Additionally, tests will be conducted to optimize the number of inference iterations.

3.1.4 Experiment 2: Direct Repurposed Diffusion Model

In this experiment, the selected diffusion model will be trained on medical CT images and then directly applied to denoise synchrotron LDCT images at various dose levels. Additionally, tests will be conducted to determine the optimal number of inference iterations.

3.1.5 Experiment 3: Fine-Tuned Repurposed Model

This experiment involves fine-tuning a diffusion model, initially trained on medical CT images, using the train set from the synchrotron CT dataset. The model will then be applied

to denoise the synchrotron LDCT test set data across various dose levels. The objective is to refine the model's weights through fine-tuning and evaluate whether this domain adaptation technique improves its performance. Additionally, tests will be conducted to determine the optimal number of inference iterations.

3.2 Datasets

This section presents the datasets available for use in the development of the proposed methodology. Sections 3.2.1, 3.2.2, 3.2.3 and 3.2.4 presents medical CT datasets and Section 3.2.5 a synchrotron CT dataset.

3.2.1 2016 Low Dose CT Grand Challenge Dataset

The 2016 Low Dose CT Grand Challenge [39] was a competition sponsored by the National Institute of Health (NIH), the American Association of Physicists in Medicine (APPM) and the Mayo Clinic with primary goal to evaluate and compare denoising and iterative reconstruction techniques for improving liver lesion detection in low-dose CT scans. In 2021, the complete competition dataset has been made available and integrated into the Low Dose CT Image and Projection Dataset presented in Section 3.2.2. A document making possible to match the samples from the Grand Challenge within the Low Dose CT Image and Projection Dataset is available as the coding use for the patient cases on both datasets vary among them.

The dataset is composed of projection and image data of contrast-enhanced abdominal CT from 30 anonymized patient cases, where 10 were initially made available to contestants as a train set and 20 were used as test set. For the 10 train set cases, normal and quarter dose versions of the data were provided while for the 20 test sets cases only the quarter dose data was made available. The full-dose data were acquired at 120kV with 200 effective Quality Reference Milliampere-seconds (QRM) while quarter dose data, corresponding to 120kv and 50 QRM, were simulated by adding noise to the full-dose data with the technique proposed by Yu et al. [58]. For all cases, the image data reconstruction was done from the projection data by applying a FBP technique. Four different reconstructions were made, two using the B30 reconstruction kernel but with slice thickness of 1mm and 3mm, and the other two used the same slice thickness variation but used the D45 reconstruction kernel.

Among with the previously described data, simulated phantom data was also made available following the same full-dose, quarter-dose and reconstructions parameters from patient cases data. Except for the 50 QRM phantom data, all the rest of the phantom data was simulated.

3.2.2 Low Dose CT Image and Projection Dataset

The Low Dose CT Image and Projection Dataset [11, 38] gathers a total of 299 patients CT exams data whose 99 of them are head scans, 100 are chest scans and 100 are abdomen scans. The images within this dataset are half part from a SOMATOM Definition Flash CT scanner from Siemens (150 cases) and the other half from a Lightspeed VCT CT scanner from General Electric (GE) (149 cases). The data provided for all cases were: Routine dose and simulated low-dose projection data, reconstructed image data for the full-dose cases and also clinical data for identifying patient pathologies identifiable from the data.

The acquisition of the dataset was done using routine dose levels according to standard clinical protocols and the anatomical body region. A low-dose version of all data was simulated by applying the technique proposed by Yu et al. [58], which models the impact of tube current modulation, the bow-tie filter and electronic noise. For head and abdomen cases the low-dose version provided, simulates images equivalent to exposing the patient to 25% of the routine dose while for chest cases it was equivalent to 10% of the routine dose. The full-dose reconstructed images were acquired by applying a FBP method. Besides that, clinical data for identifying patient pathologies related to the projection and image data was also provided.

Finally, all data is available in the open and vendor-neutral format DICOM-CT-PD [10] and can be found on The Cancer Imaging Archive (TCIA) [38].

3.2.3 LIDC-IDRI Dataset

The Lung Image Database Consortium image collection (LIDC-IDRI) [1] consist of diagnostic and lung cancer screening thoracic CT scans with marked-up annotated lesions. It contains 1018 helical thoracic CT scans of 1010 patients and was created by seven academic center and eight medical imaging companies.

Each case is composed by a clinical thoracic CT scan with a Extensible Markup Language (XML) file with annotations done by four experienced thoracic radiologist about 3 categories of nodules. Both standard-dose scans and lower-dose scans are provided in the dataset but projection data was not made available [34].

Finally, all data is available in the open and vendor-neutral format DICOM-CT-PD [10] and can be found on TCIA [1].

3.2.4 LoDoPaB-CT Dataset

The Low-Dose Parallel Beam X-Ray Computed Tomography (LoDoPaB-CT) [34] dataset is a manual selection of 812 patient cases from the LIDC-IDRI database presented in Section 3.2.3 which scans considered "too noisy" were removed.

The dataset goal is to be a benchmark for evaluating reconstruction methods but it can also be leveraged for denoising algorithm evaluation. The dataset is split in training, validation and testing parts, containing 35820 images, 3522 images and 3553 images, respectively.

As projection data is not provided in the LIDC-IDRI dataset, the creators of LoDoPaB-CT simulated the projection data from the reconstructed images. The simulation procedure involves preprocessing the CT images by cropping, dequantizing, and normalizing them. The images are then upscaled and subjected to a mimic parallel beam to simulate low-dose CT measurements. Poisson-distributed noise is added to mimic real-world conditions. To avoid photon starvation, areas with zero photon counts are replaced by a counting of 0.1.

3.2.5 X-Ray Tomography Data Bank

The X-Ray Tomography Data Bank also named as just TomoBank [14] is a public repository currently maintained in the Argonne Leadership Computing Facility from the Argonne National Laboratory at USA. The repository contains experimental data from different types of real or simulated samples from different synchrotron facilities. Together with the data,

a python script for reading the data and performing a simple tomographic reconstruction is also provided.

3.3 Evaluation Metrics

This section presents commonly used metrics for quantifying the enhancement promoted by a denoising algorithm over noisy images. Some of them consider the global information of the image while others may be applied in a given Region of Interest (ROI) within the image.

3.3.1 Root Mean Square Error (RMSE)

The Root Mean Square Error (RMSE) [20] is a metric vastly used in many different applications to measure the error between an observed value and an estimation of it. In an image denoising application, its calculation is done between the noise-free image and a denoised estimation generated by an algorithm.

The calculation of the RMSE is done by taking the square root of the Mean Square Error (MSE), which is the average difference between each pixel from the noise-free image versus the denoised estimation. Equations 3.1 and 3.2 defines how the MSE and Root Mean Square Error (RMSE) are calculated, where I_{ij} and $\hat{I}_{ij}$ are the pixel value at position (i, j) in the noise-free image and denoised image, while M and N defines both images height and width values.

$$\text{MSE} = \frac{1}{M \times N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \left(I_{ij} - \hat{I}_{ij} \right)^2 \tag{3.1}$$

$$\text{RMSE} = \sqrt{\text{MSE}} \tag{3.2}$$

Low RMSE values indicates better denoising performance of an algorithm as the difference between the noise-free and denoised image would be small which indicates the algorithm capability to approximate the denoised image to the ideal reference.

3.3.2 Peak Signal-to-Noise Ratio (PSNR)

The Peak Signal-to-Noise Ratio (PSNR) [25] metric is a commonly used metric for assessing the quality of image reconstructions, compression and denoising algorithms. It is commonly expressed in Decibels (dB) due to wide dynamic range of signals.

Its calculation is done based on the MSE between an ideal reference image I and the approximation image $\hat{I}$ obtained by the applied algorithm. Equation 3.3 defines how PSNR is calculated where L_I is the maximum possible intensity value of a pixel from image I, the MSE is calculated as defined in Equation 3.1, being M and N the dimensions of images I and $\hat{I}$, and the logarithmic base is applied to express the metric in dB. Minimizing the MSE is desirable as it means that both images I and $\hat{I}$ would be very similar. Consequently, the higher the PSNR value the better the achieved result.

Into a denoising application, the image I is a noise-free referential image and $\hat{I}$ the resulting denoised image acquired by applying the denoising algorithm over the noisy version of image I.

$$\text{PSNR} = 10 \times \log_{10} \left(\frac{L_I^2}{\text{MSE}} \right) \tag{3.3}$$

3.3.3 Structural Similarity Index Measure (SSIM)

The Structural Similarity Index Measure (SSIM) [52] is a perception-based metric for measuring the similarity between two images, different from RMSE and PSNR that focus on calculating absolute errors. Its calculation is done based on the degradation of structural information from two windows $\mathbf{x}$ and $\mathbf{y}$ of common size $M \times N$ extracted from the same position of both images being compared.

It evaluates three images aspects between both regions: Luminance ($l(\mathbf{x}, \mathbf{y})$) defined in Equation 3.4, contrast ($c(\mathbf{x}, \mathbf{y})$) defined in Equation 3.5 and structure ($s(\mathbf{x}, \mathbf{y})$) defined in Equation 3.6, where μ, σ^2, σ are the mean pixel intensity, the variance and standard deviation from a given region and σ_{xy} the covariance of regions $\mathbf{x}$ and $\mathbf{y}$. Additionally, C_1, C_2 and C_3 are stabilization variables defined in Equations 3.7, 3.8, and 3.9, where L is the dynamic range of the pixel values (e.g, 255 for 8-bit grayscale images), and k_1 and k_2 are small constants $\ll 1$ [52]. Wang et al. [52] used the arbitrary values of $k_1 = 0.01$ and $k_2 = 0.03$ in their SSIM proposal paper and comment "fairly insensitive" variations of SSIM to variations of those values.

$$l(\mathbf{x}, \mathbf{y}) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \tag{3.4}$$

$$c(\mathbf{x}, \mathbf{y}) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \tag{3.5}$$

$$s(\mathbf{x}, \mathbf{y}) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \tag{3.6}$$

$$C_1 = (k_1 L)^2 \tag{3.7}$$

$$C_2 = (k_2 L)^2 \tag{3.8}$$

$$C_3 = \frac{C_2}{2} \tag{3.9}$$

The SSIM is them calculated as a weighted combination of the luminance, contrast and structure comparisons between regions $\mathbf{x}$ and $\mathbf{y}$ as defined in Equation 3.10, where α, β and γ coefficients defines the weights given for each aspect. If the same importance is considered for all three aspects ($\alpha = \beta = \gamma = 1$), Equation 3.10 can be simplified in Equation 3.11.

$$\text{SSIM}(\mathbf{x}, \mathbf{y}) = [l(\mathbf{x}, \mathbf{y})]^\alpha \times [c(\mathbf{x}, \mathbf{y})]^\beta \times [s(\mathbf{x}, \mathbf{y})]^\gamma \tag{3.10}$$

$$\text{SSIM}(\mathbf{x}, \mathbf{y}) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{3.11}$$

Wang et al. [52] proposed that in order to calculate a single overall quality measure of the entire image, it is possible to calculate the Mean Structural Similarity Index Measure (MSSIM) by calculating the the SSIM over M different windows within image and taking the mean value of it. A weighted average can also be calculate if changes are done on the α, β and γ coefficients. A multi-scale extension of SSIM called Multiscale Structural Similarity Index Measure (MS-SSIM) was also proposed by Wang, Simoncelli and Bovik [53].

Finally, SSIM value varies between -1 and 1 where the nearest the acquired value is from 1, the more similar both compared images are.

3.4 Computational Resources

The project will be developed on Python programming language due to the vast number of packages available for developing ML, ANN and image processing solutions.

The experiments will be executed on a personal computer, Google Colab [22] or a computational cluster, depending on the storage and processing requirements for developing the proposed solution.

Chapter 4

Working Plan and Execution Timeline

The working plan is composed by the following tasks and their planned execution timeline, divided by year and quarter, is presented in Table 4.1.

A. Literature review;

B. Proposal writing and realization of the qualification exam;

C. Dataset preparation;

D. Baseline metric calculation;

E. Experiments execution and results comparison;

F. Results documentation and publishing;

G. Dissertation writing and defense.

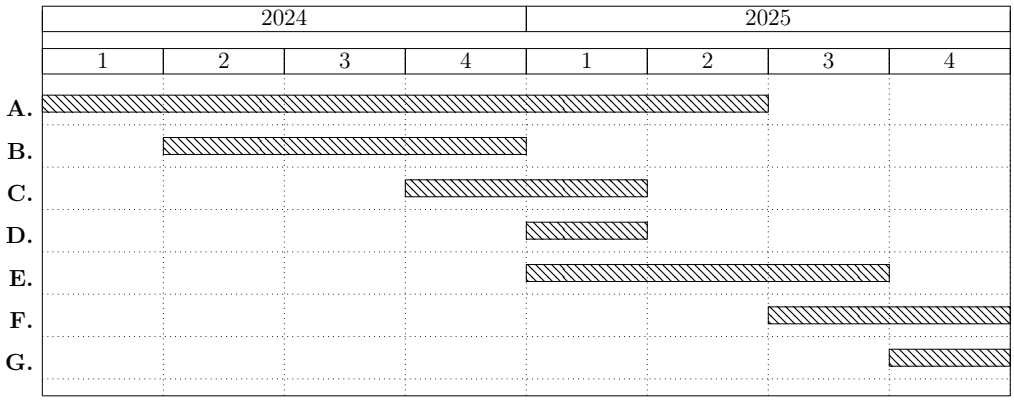


Table 4.1: Execution timeline divided by year and quarter.

References

[1] S. G. Armato III, G. McLennan, L. Bidaut, M. F. McNitt-Gray, C. R. Meyer, A. P. Reeves, B. Zhao, D. R. Aberle, C. I. Henschke, E. A. Hoffman, E. A. Kazerooni, H. MacMahon, E. J. R. Van Beek, D. Yankelevitz, A. M. Biancardi, P. H. Bland, M. S. Brown, R. M. Engelmann, G. E. Laderach, D. Max, R. C. Pais, D. P. Y. Qing, R. Y. Roberts, A. R. Smith, A. Starkey, P. Batra, P. Caligiuri, A. Farooqi, G. W. Gladish, C. M. Jude, R. F. Munden, I. Petkovska, L. E. Quint, L. H. Schwartz, B. Sundaram, L. E. Dodd, C. Fenimore, D. Gur, N. Petrick, J. Freymann, J. Kirby, B. Hughes, A. V. Casteele, S. Gupte, M. Sallam, M. D. Heath, M. H. Kuhn, E. Dharaiya, R. Burns, D. S. Fryd, M. Salganicoff, V. Anand, U. Shreter, S. Vastagh, B. Y. Croft, and L. P. Clarke. Data from The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A completed reference database of lung nodules on CT scans, 2015. The Cancer Imaging Archive. `https://doi.org/10.7937/K9/TCIA.2015.LO9QL9SX`. 3, 28

[2] A. Bansal, E. Borgnia, H.-M. Chu, J. Li, H. Kazemi, F. Huang, M. Goldblum, J. Geiping, and T. Goldstein. Cold Diffusion: Inverting Arbitrary Image Transforms Without Noise. In *Advances in Neural Information Processing Systems 36 (NeurIPS)*, volume 36, pages 41259–41282, 2023. 18, 22

[3] D. Bhatt, C. Patel, H. Talsania, J. Patel, R. Vaghela, S. Pandya, K. Modi, and H. Ghayvat. CNN Variants for Computer Vision: History, Architecture, Application, Challenges and Future Scope. *Electronics*, 10(20):2470, 2021. 12

[4] C. M. Bishop. Latent Variable Models. In *Learning in Graphical Models*, pages 371–403. Springer, 1998. 13

[5] C. M. Bishop and H. Bishop. *Deep Learning: Foundations and Concepts*. Springer Nature, 2023. 11, 12, 13, 14, 15, 16, 17, 18

[6] F. E. Boas and D. Fleischmann. CT artifacts: causes and reduction techniques. *Imaging in Medicine*, 4(2):229–240, 2012. 2

[7] U. Bonse and F. Busch. X-Ray computed microtomography (μCT) using synchrotron radiation (SR). *Progress in Biophysics and Molecular Biology*, 65(1):133–169, 1996. 1

[8] A. Buades, B. Coll, and J.-M. Morel. A non-local algorithm for image denoising. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 60–65 vol. 2, 2005. 3, 19, 26

[9] J. T. Bushberg, J. A. Seibert, J. Leidholdt, Edwin M, and J. M. Boone. *The essential physics of medical imaging*. Lippincott Williams & Wilkins, 2011. 1, 2, 5, 6, 7, 8, 9, 10, 11

[10] B. Chen, X. Duan, Z. Yu, S. Leng, L. Yu, and C. McCollough. Development and Validation of an Open Data Format for CT Projection Data. *Medical Physics*, 42(12):6964–6972, 2015. 28

[11] B. Chen, S. Leng, L. Yu, D. H. III, J. Fletcher, and C. McCollough. An open library of CT patient projection data. In *Medical Imaging 2016: Physics of Medical Imaging*, volume 9783, page 97831B. International Society for Optics and Photonics, SPIE, 2016. 3, 27

[12] H. Chen, Y. Zhang, M. K. Kalra, F. Lin, Y. Chen, P. Liao, J. Zhou, and G. Wang. Low-Dose CT With a Residual Encoder-Decoder Convolutional Neural Network. *IEEE Transactions on Medical Imaging*, 36(12):2524–2535, 2017. 18, 19, 20

[13] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering. *IEEE Transactions on Image Processing*, 16(8):2080–2095, 2007. 2, 19, 26

[14] F. De Carlo, D. Gürsoy, D. J. Ching, K. J. Batenburg, W. Ludwig, L. Mancini, F. Marone, R. Mokso, D. M. Pelt, J. Sijbers, and M. Rivers. TomoBank: a tomographic data repository for computational X-Ray science. *Measurement Science and Technology*, 29(3):034004, 2018. 3, 28

[15] X. Duan, X. F. Ding, N. Li, F.-X. Wu, X. Chen, and N. Zhu. Sparse2Noise: Low-dose synchrotron X-Ray tomography without high-quality reference data. *Computers in Biology and Medicine*, 165:107473, 2023. 1, 2

[16] X. Duan, N. Li, X. Chen, and N. Zhu. Characterization of Tissue Scaffolds Using Synchrotron Radiation Microcomputed Tomography Imaging. *Tissue Engineering Part C: Methods*, 27(11):573–588, 2021. 1

[17] Q. Gao, Z. Li, J. Zhang, Y. Zhang, and H. Shan. CoreDiff: Contextual Error-Modulated Generalized Diffusion Model for Low-Dose CT Denoising and Generalization. *IEEE Transactions on Medical Imaging*, 43(2):745–759, 2024. 1, 2, 3, 22, 23, 24

[18] Q. Gao and H. Shan. CoCoDiff: a contextual conditional diffusion model for low-dose CT image denoising. In *Developments in X-Ray Tomography XIV*, volume 12242, page 122420I. International Society for Optics and Photonics, SPIE, 2022. 1, 2, 3, 21, 22

[19] K. Gong, K. Johnson, G. El Fakhri, Q. Li, and T. Pan. PET image denoising based on denoising diffusion probabilistic model. *European Journal of Nuclear Medicine and Molecular Imaging*, 51(2):358–368, 2024. 3

[20] R. Gonzalez and R. Woods. *Digital Image Processing*. Prentice Hall, 2008. 10, 11, 29

[21] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative Adversarial Nets. *Advances in Neural Information Processing Systems 27 (NeurIPS)*, 27, 2014. 14

[22] Google Colab. Goggle Colaboratory. *https://colab.google/*, 2024. Online; accessed on August 6, 2024. 31

[23] C. A. Hamm, H. Mallison, O. Hampe, D. Schwarz, J. Mews, J. Blobel, A. S. Issever, and P. Asbach. Efficiency, workflow and image quality of clinical computed tomography scanning compared to photogrammetry on the example of a *Tyrannosaurus rex* skull from the Maastrichtian of Montana, USA. *Journal of Paleontological Techniques*, 21:1–13, 2018. 1

[24] J. Ho, A. Jain, and P. Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems 33 (NeurIPS)*, 33:6840–6851, 2020. 2, 17

[25] A. Hore and D. Ziou. Image Quality Metrics: PSNR vs. SSIM. In *20th International Conference on Pattern Recognition*, pages 2366–2369. IEEE, 2010. 29

[26] D. Hu, Y. K. Tao, and I. Oguz. Unsupervised denoising of retinal OCT with diffusion probabilistic model. In *Medical Imaging 2022: Image Processing*, volume 12032, pages 25–34. SPIE, 2022. 3

[27] Z. Huang, J. Zhang, Y. Zhang, and H. Shan. DU-GAN: Generative Adversarial Networks With Dual-Domain U-Net-Based Discriminators for Low-Dose CT Denoising. *IEEE Transactions on Instrumentation and Measurement*, 71:1–12, 2022. 20, 21

[28] V.-P. Karjalainen, M. A. Finnilä, P. L. Salmon, and S. Lipkin. Micro-computed tomography imaging and segmentation of the archaeological textiles from Valmarinniemi. *Journal of Archaeological Science*, 160:105871, 2023. 1

[29] A. Kazerouni, E. K. Aghdam, M. Heidari, R. Azad, M. Fayyaz, I. Hacihaliloglu, and D. Merhof. Diffusion models in medical imaging: A comprehensive survey. *Medical Image Analysis*, 88:102846, 2023. 2

[30] D. Kingma, T. Salimans, B. Poole, and J. Ho. Variational Diffusion Models. *Advances in Neural Information Processing Systems 34 (NeurIPS)*, 34:21696–21707, 2021. 18

[31] K. S. H. Kulathilake, N. A. Abdullah, A. Q. M. Sabri, and K. W. Lai. A review on deep learning approaches for low-dose computed tomography restoration. *Complex & Intelligent Systems*, 9(3):2713–2745, 2023. 1, 2

[32] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015. 2

[33] LeCun, Yann and Kavukcuoglu, Koray and Farabet, Clement. Convolutional Networks and Applications in Vision. In *International Symposium on Circuits and Systems*, pages 253–256. IEEE, 2010. 12

[34] J. Leuschner, M. Schmidt, D. O. Baguer, and P. Maass. LoDoPaB-CT, a benchmark dataset for low-dose computed tomography reconstruction. *Scientific Data*, 8(1):109, 2021. 3, 28

[35] X. Liu, Y. Xie, S. Diao, S. Tan, and X. Liang. A diffusion probabilistic prior for zero-shot low-dose CT image denoising. *arXiv preprint arXiv:2305.15887*, 2023. 2, 3, 22

[36] Z. Liu, T. Bicer, R. Kettimuthu, D. Gursoy, F. D. Carlo, and I. Foster. TomoGAN: low-dose synchrotron X-Ray tomography with generative adversarial networks: discussion. *Journal of the Optical Society of America A*, 37(3):422–434, 2020. 1, 2, 9, 10, 18

[37] Martin Arjovsky and Soumith Chintala and Léon Bottou. Wasserstein Generative Adversarial Networks. In *International Conference on Machine Learning (ICML)*, volume 70 of *Proceedings of Machine Learning Research*, pages 214–223, 2017. 16

[38] C. McCollough, B. Chen, D. R. Holmes III, X. Duan, Z. Yu, L. Yu, S. Leng, and J. Fletcher. Low Dose CT Image and Projection Data, 2020. The Cancer Imaging Archive. `https://doi.org/10.7937/9NPB-2637`. 27, 28

[39] C. H. McCollough, A. C. Bartley, R. E. Carter, B. Chen, T. A. Drees, P. Edwards, D. R. Holmes III, A. E. Huang, F. Khan, S. Leng, K. L. McMillan, G. J. Michalak, K. M. Nunez, L. Yu, and J. G. Fletcher. Low-dose CT for the Detection and Classification of Metastatic Liver Lesions: Results of the 2016 Low Dose CT Grand Challenge. *Medical Physics*, 44(10):e339–e352, 2017. 3, 27

[40] E. Okuno and E. M. Yoshimura. *Física das Radiações*. Oficina de Textos, 2016. 1, 9

[41] F. Rosenblatt. *Principles of neurodynamics: Perceptrons and the theory of brain mechanisms*, volume 55. Spartan books Washington, D.C., 1962. 12

[42] S. V. M. Sagheer and S. N. George. A review on medical image denoising algorithms. *Biomedical Signal Processing and Control*, 61:102036, 2020. 2

[43] P. Savoia, S. K. Jayanthi, and M. C. Chammas. Focused Assessment with Sonography for Trauma (FAST). *Journal of Medical Ultrasound*, 31(2):101–106, 2023. 1

[44] R. Schilling, B. Jastram, O. Wings, D. Schwarz-Wings, and A. S. Issever. Reviving the dinosaur: virtual reconstruction and three-dimensional printing of a dinosaur vertebra. *Radiology*, 270(3):864–871, 2014. 1

[45] K. Simonyan and A. Zisserman. Very Deep Convolutional Networks for Large-scale Image Recognition. *arXiv preprint arXiv:1409.1556*, 2014. 19

[46] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, and S. Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International Conference on Machine Learning (ICML)*, volume 37 of *Proceedings of Machine Learning Research*, pages 2256–2265. PMLR, 2015. 17

[47] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020. 2, 21

[48] Y. R. Tonin. Coherent X-Ray Diffraction Imaging: Image reconstruction via a matrix model of the inhomogenous Helmholtz equation. Master's thesis, Unicamp, 2022. 7

[49] N. T. Trung, D.-H. Trinh, N. L. Trung, and M. Luong. Low-dose CT image denoising using deep convolutional neural networks with extended receptive fields. *Signal, Image and Video Processing*, 16(7):1963–1971, 2022. 1, 2

[50] L. Vásárhelyi, Z. Kónya, A. Kukovecz, and R. Vajtai. Microcomputed tomography–based characterization of advanced materials: a review. *Materials Today Advances*, 8:100084, 2020. 1

[51] T. Wang, Y. Lei, Z. Tian, X. Dong, Y. Liu, X. Jiang, W. J. Curran, T. Liu, H.-K. Shu, and X. Yang. Deep learning-based image quality improvement for low-dose computed tomography simulation in radiation therapy. *Journal of Medical Imaging*, 6(4):043504–043504, 2019. 1

[52] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 30

[53] Z. Wang, E. P. Simoncelli, and A. C. Bovik. Multiscale Structural Similarity for Image Quality Assessment. In *Asilomar Conference on Signals, Systems & Computers (ACSSC)*, volume 2, pages 1398–1402. IEEE, 2003. 30

[54] L. Weng. What are diffusion models? *https://lilianweng.github.io/posts/2021-07-11-diffusion-models*, 2021. Online; accessed on July 8, 2024. 14

[55] P. Willmott. *An introduction to synchrotron radiation: techniques and applications*. John Wiley & Sons, 2019. 7, 8

[56] W. Xia, Q. Lyu, and G. Wang. Low-Dose CT Using Denoising Diffusion Probabilistic Model for 20× Speedup. *arXiv preprint arXiv:2209.15136*, 2022. 3, 21

[57] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, Y. Zhang, L. Sun, and G. Wang. Low-Dose CT Image Denoising Using a Generative Adversarial Network With Wasserstein Distance and Perceptual Loss. *IEEE Transactions on Medical Imaging*, 37(6):1348–1357, 2018. 19, 20

[58] L. Yu, M. Shiung, D. Jondal, and C. H. McCollough. Development and Validation of a Practical Lower-Dose-Simulation Tool for Optimizing Computed Tomography Scan Protocols. *Journal of Computer Assisted Tomography*, 36(4):477–487, 2012. 26, 27, 28