ENHANCING QUALITY OF LOW-DOSE CT SCANS VIA GENERATIVE DIFFUSION MODELS

ENHANCING QUALITY OF LOW-DOSE CT SCANS VIA GENERATIVE DIFFUSION MODELS

By Seyed Mohammad Mehdi Hassani Najafbadi,

A Thesis Submitted to the School of Graduate Studies in the Partial Fulfillment of the Requirements for the Degree Master of Applied Science

McMaster University

Master of Applied Science (2024)

Hamilton, Ontario (ECE Department)

TITLE: ENHANCING QUALITY OF LOW-DOSE CT SCANS VIA GENERATIVE DIFFUSION MODELS

AUTHOR: Seyed Mohammad Mehdi Hassani Najafbadi

B.Sc., (Computer Engineering)

McMaster University, Hamilton, Canada

SUPERVISOR: Dr. Shahram Shirani

NUMBER OF PAGES: xiii, 70

Abstract

The enduring challenge in computed tomography (CT) imaging is mitigating the radiation risks associated with high-dose protocols while maintaining image quality. This research introduces an innovative approach that diverges significantly from conventional methodologies, using Generative Diffusion Models (GDM) to enhance the quality of low-dose CT scans to that of high-dose scans. This advancement is particularly pivotal as it addresses the crucial balance between minimizing radiation risk and preserving diagnostic integrity. At the heart of our approach is a distinctive application of a Convolutional Neural Network (CNN) designed not to filter noise but to meticulously identify and segregate intrinsic noise features within paired high and low-dose CT images. This method stands in contrast to traditional techniques that often rely on generic random noise models, lacking specificity to actual imaging conditions. By accurately modeling the unique noise profile of low-dose scans, we enable our GDM to undertake a reverse diffusion process, effectively reducing noise and enhancing image clarity to equal high-dose standards. The significance of transitioning from low-dose to high-dose imaging quality without additional radiation is offering a path to safer imaging protocols that do not compromise quality. We present preliminary findings substantiated by both PSNR and SSIM metrics, demonstrating improvement in image quality through our method. In addition to delineating our approach, this research draws comparisons with existing methods, particularly focusing on PALLETE, a known algorithm in the field. Our comparative analysis illustrates the superiority of our model in terms of image quality, showcasing our method's potential for enhancement in radiological imaging.

Acknowledgements

First and foremost, I would like to express my heartfelt gratitude to my supervisor, Dr. Shahram Shirani, for his unwavering guidance and generous support throughout this journey. I am truly delighted to work under his supervision, and I have learned a lot from him, not only technically but also in terms of his exemplary conduct and interactions. Without his profound insights, continuous support, and encouragement, this work would never have come to fruition

I extend my sincere appreciation to the administrative staff, especially, Cheryl Gies, for their efficient coordination and assistance in navigating the intricacies of this academic pursuit.

Abbreviations

CNN	Convolution Neural Network
CT	Computed Tomography
GDM	Generative Diffusion Models
PSNR	Peak Signal-to-Noise Ratio
SSIM	Structural Similarity Index Measure
IR	Iterative Reconstruction
FBP	Filtered Back Projection
GANs	Generative Adversarial Networks
ALARA	Low As Reasonably Achievable
NCDM	Noise-Characterized Diffusion Model
LDCT	Low-Dose CT images
HDCT	High-dose CT images

Contents

List of Figures

List of Tables

Declaration of Authorship

I, Seyed Mohammad Mehdi Hassani Najafbadi, declare that this thesis titled, “ENHANCING QUALITY OF LOW-DOSE CT SCANS VIA GENERATIVE DIFFUSION MODELS” and the work presented in it are my own.

Chapter 1

Introduction

The advent of Computed Tomography (CT) has undeniably revolutionized the medical field, offering unparalleled internal views for diagnostic accuracy. However, the use of ionizing radiation in CT scanning has raised significant health concerns, primarily due to the potential risk of cancer and other radiation-induced conditions (Brenner and Hall 2007; Pearce et al. 2012). These concerns have propelled a shift towards low-dose CT protocols, aimed at minimizing patient exposure to harmful radiation. While these protocols are a step forward in enhancing patient safety, they introduce a new challenge: increased image noise, which can significantly compromise diagnostic clarity and reliability (Pugliesi 2018). The balance between reducing radiation exposure and maintaining high-quality diagnostic images presents a critical conundrum in medical imaging, highlighting the need for innovative solutions. Traditional approaches to noise reduction in CT imaging, such as Filtered Back Projection (FBP) and Iterative Reconstruction (IR), have provided pathways to clearer images at lower doses. However, these techniques often fall short when balancing noise suppression against detail preservation, sometimes introducing artifacts or losing vital diagnostic information (Willemink and Noël 2019). As the limitations of these conventional methods became apparent, the medical imaging community turned its attention to more advanced solutions. In recent years, artificial intelligence (AI), particularly deep learning, has emerged as a powerful tool in various domains, including medical imaging. Convolutional Neural Networks (CNNs), a class of deep learning models, have shown particular promise in enhancing image quality by learning complex noise patterns and structures directly from the data (Wang et al. 2019). These models have redefined expectations for what is possible in terms of noise reduction and detail enhancement, but they are not without their drawbacks. Generalization across different noise levels and patient scenarios remains a challenge, as does the computational intensity of training and deploying these models (Rueckert and Schnabel 2019).

Generative Adversarial Networks (GANs) introduced a novel paradigm by generating detailed, high-quality images from low-dose CT data. Despite their potential, GANs are often criticized for their tendency towards mode collapse and the introduction of non-authentic details, which could mislead diagnosis (Yi et al. 2019).

Enter Generative Diffusion Models (GDM), a new class of generative models that have taken the field of image processing by storm. Known for their impressive ability to generate and enhance images, GDMs offer a structured approach to image synthesis, gradually transforming noise into detailed, coherent structures through a controlled process. This method holds significant promise for medical imaging, particularly for enhancing low-dose CT scans, by learning and reversing the noise distribution inherent to these images (Hung et al. 2023).

Despite the promise of GDMs, their application in enhancing low-dose CT images remains underexplored. The majority of existing studies have concentrated on natural image processing, with only a handful venturing into the complexities of medical imaging and even fewer addressing the specific challenges of CT noise. Our research aims to bridge this gap, leveraging the unique capabilities of GDMs in conjunction with CNNs to target the distinctive noise patterns of low-dose CT images. By doing so, we propose a novel solution that not only adheres to the ALARA(Low As Reasonably Achievable) principle (Miller and Schauer 1983) but also pushes the boundaries of what's achievable in low-dose CT imaging quality.

1.1 Background and Context

1.1.1 Historical Development of CT Imaging Technology

The invention of Computed Tomography (CT) imaging in the early 1970s by Sir Godfrey Hounsfield and Dr. Allan Cormack marked a revolutionary advancement in medical diagnostics (Hounsfield 1973; Cormack 1973). This innovative technology, which garnered them the Nobel Prize in Medicine in 1979, employed X-ray measurements from various angles to create cross-sectional images of the body, introducing a groundbreaking perspective in medical examinations.

Initially, CT scanners were exclusively used for head imaging, requiring extensive hours for data acquisition and image reconstruction. However, recognizing the significant potential of CT scans, technological advancements quickly ensued. The development of whole-body scanners by the late 1970s substantially broadened the scope of CT imaging applications (Ambrose 1976).

Technological progress continued to enhance the functionality of CT scanners, notably with the introduction of spiral (or helical) CT technology in the 1980s, which allowed for continuous data acquisition as patients moved through the scanner (Kalender et al. 1989). This innovation dramatically reduced scan times and enhanced image resolution, leading to more detailed and precise diagnoses. The late 1990s saw further advancements with the development of multislice (or multidetector) CT scanners, which significantly decreased scanning times and improved image quality by capturing multiple slices in a single rotation, facilitating comprehensive three-dimensional reconstructions (Flohr et al. 2005).

CT imaging has become a fundamental component of medical diagnostics and treatment planning, providing intricate details of bones, blood vessels, and soft tissues. It has proven invaluable in diagnosing a variety of conditions, such as injuries, infections, tumors, and vascular diseases (Smith and Webb 2010). The ability of CT imaging to deliver rapid and accurate diagnostic information is especially critical in emergencies, where it can be life-saving (*Body CT (CAT Scan)* 2024).

Moreover, CT imaging is instrumental in the planning and execution of medical procedures and surgeries, evaluating the effectiveness of treatments, and monitoring disease progression (Edelmers et al. 2024). Its extensive application across various medical fields, including oncology, cardiovascular medicine, neurology, and orthopedics, underscores its importance in modern healthcare.

1.1.2 Evolution from Traditional CT Scanning to Low-Dose CT Protocols

Despite the substantial benefits of CT imaging, traditional scanning methods have been scrutinized due to the associated health risks from ionizing radiation, such as increased cancer risk (Brenner and Hall 2007; Pearce et al. 2012). This concern has led to the adoption of low-dose CT protocols, emphasizing the ALARA principle to minimize radiation exposure while retaining image quality (Schauer and Linton 2009).

The transition to low-dose CT protocols involved multiple strategies, including adjustments in the X-ray tube current, the use of advanced noise-reduction software, and the optimization of scanning parameters to specific diagnostic needs (Hara et al. 2009). The advent of iterative reconstruction techniques and the integration of AI and machine learning into medical imaging has further advanced low-dose CT, allowing for significant reductions in radiation dosage while maintaining, or even improving, the quality of diagnostic images (Beister et al. 2012; Li et al. 2014).

As CT technology continues to advance, it is expected to uphold the delicate balance between patient safety and diagnostic efficacy, ensuring CT imaging remains a vital tool in medical diagnostics while addressing radiation safety concerns (Dobbins et al. 2014).

1.2 Advances in CT Imaging: Traditional Techniques, AI Integration, and Future Direction

Noise reduction in CT imaging has significantly evolved, transitioning from basic filtering techniques to more sophisticated algorithms aimed at improving image clarity and diagnostic accuracy. Traditional methods like Filtered Back Projection (FBP) have been a mainstay due to their computational efficiency, crucial for time-sensitive applications such as emergency diagnostics. However, the limitation of FBP lies in its predisposition to noise, particularly under low-dose conditions, which can obscure critical details in the images (Hsieh 2003; Karimi et al. 2016).

The development of Iterative Reconstruction (IR) techniques marked a significant leap forward. IR methods, such as Algebraic Reconstruction Technique (ART) and Model-Based Iterative Reconstruction (MBIR), systematically refine image quality by reducing noise and artifacts. They work by approximating a solution through iterative processing, leading to higher quality images that facilitate better diagnosis, particularly in challenging areas like oncology and cardiovascular diseases (Beister et al. 2012; Ramirez Giraldo et al. 2011). Despite the advantages, the higher computational demands of IR compared to FBP require more powerful processing hardware and can result in longer image processing times, which might limit their use in emergency settings (Li et al. 2014).

The integration of Artificial Intelligence (AI) in CT imaging has been transformative, marking a departure from conventional imaging techniques. Deep learning, a subset of AI, utilizes algorithms modeled after the human brain, enabling significant enhancements in image analysis, noise reduction, and interpretation accuracy (Greenspan et al. 2016; Wang et al. 2017). Convolutional Neural Networks (CNNs), a class of deep learning models, have demonstrated exceptional capabilities in identifying complex patterns in imaging data, enabling the detection of anomalies that may be invisible to the human eye (Litjens et al. 2017; Shen et al. 2017).

AI technologies not only provide superior noise reduction compared to traditional methods but also significantly reduce the amount of radiation needed to produce high-quality images, aligning with patient safety initiatives (Kawamura et al. 2024). Moreover, AI's predictive analytics can assist in prognosis and treatment planning, leading to more personalized patient care. However, challenges remain, including the need for large annotated data sets for training and issues related to the interpretability of AI models, known as the "black box" problem. Advances in explainable AI are beginning to address these concerns, improving trust and understanding among clinicians (Yamashita et al. 2018).

The synergistic integration of traditional noise reduction techniques and AI into CT imaging heralds a new era in medical diagnostics. This evolution not only enhances diagnostic precision but also significantly improves patient safety by minimizing radiation exposure. The future of CT imaging is likely to be characterized by further personalization, with AI-driven algorithms tailoring imaging protocols to individual patient characteristics and clinic scenarios (Pianykh 2020).

Moreover, as AI continues to evolve, we can anticipate the development of more advanced diagnostic tools that integrate real-time image analysis, offering immediate insights during imaging procedures. This could revolutionize areas such as interventional radiology and oncological surgery, where precise imaging is crucial. However, realizing the full potential of these advancements requires overcoming existing challenges, such as ethical concerns related to data privacy, standardizing AI applications in clinical practice, and ensuring equitable access to this cutting-edge technology.

In conclusion, while traditional noise reduction techniques continue to play a vital role in CT imaging, the integration of AI promises to enhance every aspect of this field, from diagnostic accuracy to patient safety.

1.3 Generative Models in Medical Imaging

1.3.1 Generative Adversarial Networks (GANs) in Medical application

The integration of generative models, especially Generative Adversarial Networks (GANs), into medical imaging represents a significant leap forward, offering innovative solutions and transformative potentials in enhancing Computed Tomography (CT) imaging and beyond. This section delves into the recent advancements, applications, and multifaceted challenges of generative models, emphasizing their growing impact on the medical imaging landscape.

Since their inception by Ian Goodfellow et al. in 2014, GANs have revolutionized the paradigm of machine learning with their unique architecture comprising two competing neural networks: the generator and the discriminator. This adversarial process has led to the generation of high-fidelity images, which are increasingly leveraged in medical imaging for their ability to produce detailed and realistic synthetic images (Goodfellow et al. 2014; Yi et al. 2019).

In the realm of CT imaging, GANs have found profound applications including the enhancement of image quality in low-dose CT scans, generating virtual contrast-enhanced images without actual contrast administration, and creating synthetic datasets for training and validation purposes (Kazeminia et al. 2020). These applications significantly contribute to dose reduction, improved patient safety, and broader training datasets for machine-learning models.

Furthermore, GANs are instrumental in addressing incomplete datasets by generating synthetic images to fill gaps, thereby enabling more comprehensive analyses. Their application extends to the generation of auxiliary images for multimodal disease diagnosis and the simulation of patient-specific anatomical models for preoperative planning and education (Costa et al. 2018).

Despite the promising advancements, GANs face critical challenges in CT imaging. The risk of generating anatomically incorrect or misleading features could lead to misdiagnoses or inappropriate treatment decisions. The 'black box' nature of these networks, coupled with the difficulty in validating synthetic images against real counterparts, underscores the need for transparent and interpretable models (Han et al. 2020; Kazeminia et al. 2020).

Additionally, GANs require large volumes of high-quality, data for training, a significant challenge given the privacy concerns and data scarcity in medical settings. Addressing these issues demands robust data governance frameworks and innovative solutions to mitigate the risks of over-fitting and data bias (Yi et al. 2019).

Looking forward, the integration of GANs with other AI techniques, such as reinforcement learning and unsupervised learning, could offer new pathways for innovation. Continuous efforts in enhancing model explainability, ethical AI practices, and cross-disciplinary collaborations are essential for leveraging GANs' full potential in improving clinic outcomes and advancing medical research (Mirsky and Lee 2021).

Generative Adversarial Networks signify an important step in medical imaging, offering unparalleled opportunities for enhancing CT scan quality. Addressing the current challenges and ethical considerations is crucial for advancing their application responsibly.

1.3.2 Generative Diffusion Model in Medical Application

Generative Diffusion Models (GDMs) have emerged as a groundbreaking class of generative models that simulate the gradual diffusion process to create or edit images, including medical images. This innovative advancement in the field of medical imaging heralds a novel methodology for generating high-quality, realistic images that can bolster diagnostic accuracy, patient care, and medical research (Sohl-Dickstein et al. 2015; Dhariwal and Nichol 2021).

Distinct from GANs, which utilize an adversarial process between the generator and discriminator networks to fabricate new images, GDMs commence by introducing noise to an image and then master the reversal of this process. Through iterative refinement, these models progressively reduce the added noise, culminating in clear, detailed images. This method has demonstrated exceptional efficacy in crafting detailed medical images, where precision and accuracy are crucial (Ho et al. 2020; Song et al. 2020).

Within the realm of CT imaging, GDMs have showcased potential across several pivotal domains. A noteworthy application includes producing high-quality images from low-dose CT scans, thereby diminishing radiation exposure for patients while either maintaining or enhancing the diagnostic quality of the images. This capability is in harmony with the continuous efforts to improve patient safety in medical imaging. Additionally, GDMs have been employed in generating synthetic medical images for training purposes, thereby widening the pool of training data without compromising patient privacy or subjecting patients to additional radiation exposure.

The capacity of GDMs to generate and manipulate images also unveils new avenues for simulating diverse disease states, offering a precious tool for medical education and the planning of complex treatments or surgeries. By generating images that accurately depict a broad spectrum of pathological conditions, these models can aid clinicians and students in better comprehending and preparing for real-life scenarios.

Notwithstanding their promising applications, GDMs confront challenges, particularly regarding the ethical utilization of synthetic data and ensuring the accuracy and reliability of the generated images. The risk of producing images that are realistic yet inaccurate in their portrayal of anatomy or pathology poses a significant concern.

In essence, Generative Diffusion Models (GDMs) offer immense potential for medical imaging through their ability to generate realistic, high-quality images that support diagnosis, treatment planning, and medical education. The ability of GDMs to produce high-fidelity images from lower-quality inputs is particularly transformative. By focusing on the application of GDMs for converting high-dose CT scans into low-dose equivalents, this thesis aims to tackle one of the most pressing concerns in medical imaging: reducing patient exposure to radiation without compromising the diagnostic quality of the images.

As this technology continues to mature, navigating the associated challenges with thoughtfulness and ethical consideration will be paramount to unlocking its full potential in enhancing patient care and propelling medical knowledge forward. The transformative capability of GDMs to create accurate and detailed images from low-dose scans underscores the importance of this research direction, promising significant advancements in medical imaging technology that prioritize both patient safety and image quality. Chapter 2 provides further discussion on Generative Diffusion Models and their implications in medical imaging.

1.4 Conclusion

The continuous advancement of computed tomography (CT) imaging reflects significant progress in medical technology. Yet, amidst these advancements, significant research gaps persist, particularly in integrating advanced artificial intelligence (AI) methodologies, such as Generative Diffusion Models (GDMs), into low-dose CT imaging. These gaps highlight the unexplored potential of advanced AI in enhancing diagnostic accuracy, refining treatment strategies, and advancing patient safety.

A notable research void is the limited application of advanced AI, especially GDMs, in the realm of low-dose CT imaging. Despite the utility of technologies like Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs), GDMs remain underexplored due to their novelty and the complexities associated with medical imaging integration. This oversight is significant, as medical imaging, particularly Low-dose CT, faces unique challenges such as noise reduction that are distinct from natural image processing.

In conclusion, while the prospects of integrating AI, in Medical imaging, are promising, addressing existing research gaps is crucial. Collaborative efforts from the research community, healthcare professionals, industry, and regulatory bodies are essential to harness the full potential of AI.

Chapter 2

Diffusion Models

2.1 Introduction to Diffusion Models

Diffusion models are an emerging and fascinating subset of generative models in the expansive field of artificial intelligence (AI) and deep learning. Originating from principles observed in statistical physics, particularly from phenomena like heat diffusion and the random movement of particles, these models have transitioned into the realm of machine learning, presenting innovative applications in image processing and beyond.

The essence of diffusion models lies in the process where an image or signal is incrementally corrupted by noise, then, intriguingly, learning to reverse this corrupting process (Yang et al. 2023). The term 'diffusion' is used to describe this gradual increase of noise, mirroring the natural dispersal of particles such as molecules in a gas (Sohl-Dickstein et al. 2015).

Mathematically, diffusion models are frequently delineated using stochastic differential equations (SDEs). They define the diffusion process as a Markov chain transitioning from clean data to a noisy state across a series of steps, each involving the addition of incremental Gaussian noise. The reverse process, conversely, is characterized by SDEs aimed at moving from a noisy condition back to the original, clean data (Ho et al. 2020).

The realm of applications for diffusion models is broad and dynamically growing, covering everything from generating authentic images and converting text to image, to producing voice audio and designing molecules. Their capacity for producing highly detailed and diverse results has solidified their position as one of the most promising fields in generative AI.

Throughout their development, diffusion models have witnessed various innovations, particularly in terms of architecture and training methodologies, enhancing efficiency, realism, and the diversity of generated outcomes. With attributes such as conditional diffusion, where generation is guided by specific inputs or contexts, their versatility and efficiency have considerably improved (Hung et al. 2023).

In the broad context of AI and deep learning, diffusion models establish a unique position, as shown in Figure 2.1. They are differentiated by their robust and principled approach to generating new data, distinguishing them from other generative techniques such as GANs and VAEs. This firm theoretical foundation, along with its impressive empirical performance, has fueled an increase in both research and practical deployments, solidifying its status as a major player in the realm of generative models.

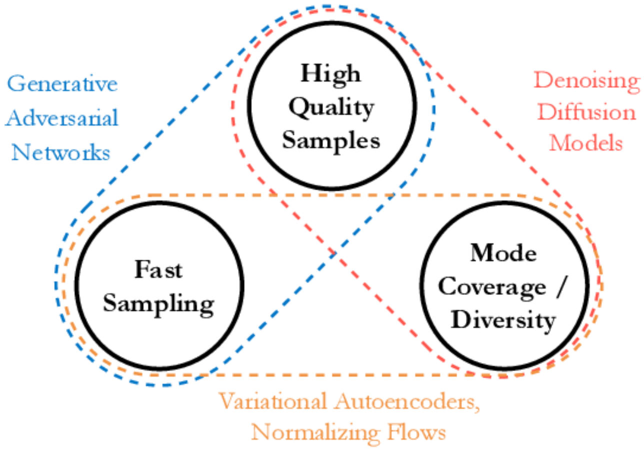

FIGURE 2.1: Generative Learning Trilemma. This diagram contrasts the capabilities of GANs, VAEs, and Diffusion models. GANs produce high-fidelity images quickly but have limited diversity. VAEs and Normalizing Flows offer more diversity but lower sample quality. Diffusion models balance diversity and quality but are slow and computationally demanding, underscoring the need for further efficiency improvements (Kazerouni et al. 2023).

2.1.1 Mathematical Framework of Diffusion Models

The core of diffusion models lies in their mathematical framework, grounded in stochastic processes. As it is described in Figure 2.2. these models operate through a two-phase process: the forward diffusion process, where noise is incrementally added to the data until only noise remains, and the reverse diffusion process, where the model learns to reconstruct the original data from the noise. Here, we describe the diffusion model through its forward and reverse processes.

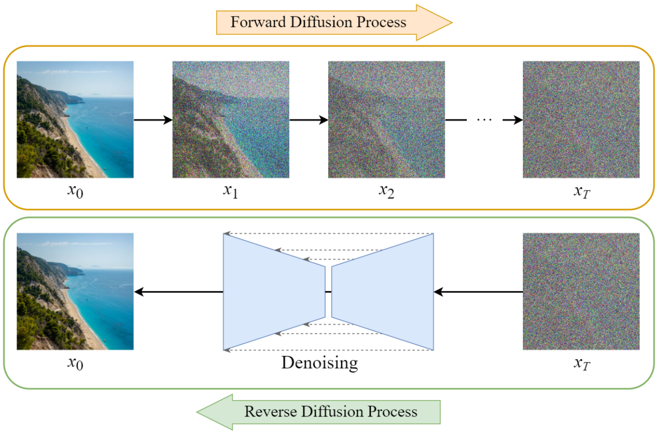

Figure 2.2: Representation of the forward and reverse diffusion process in a diffusion model (Kim and Seo 2023).

Forward Diffusion Process

The forward process, also known as the noising process, gradually transforms data into a Gaussian-distributed noise through a sequence of steps. This is mathematically represented as:

$$x_t = \sqrt{\alpha_t} x_0 + \sqrt{1 - \alpha_t}\epsilon, \quad \epsilon \sim \mathcal{N}(0, I), \quad t = 1, \ldots, T \tag{2.1}$$

In this equation, x_0 denotes the original data, x_t is the data at time step t after noise addition, α_t is a predefined variance schedule, and ϵ represents isotropic Gaussian noise.

Reverse Diffusion Process

Conversely, the reverse process, or the denoising process, aims to recover the original data from the noise. It is guided by a neural network predicting the noise added at each step of the forward process:

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{1-\alpha_t}{\sqrt{1-\alpha_t}} \epsilon_\theta(x_t, t) \right), \quad t = T, \dots, 1 \tag{2.2}$$

Here, $\epsilon_\theta(x_t, t)$ denotes the noise predicted by the neural network with parameters θ.

Objective Function

The training of diffusion models involves minimizing the difference between the actual noise added in the forward process and the predicted noise in the reverse process. This is captured by the following objective function:

$$\mathcal{L}(\theta) = \mathbb{E}_{x_0, \epsilon, t} \left[\|\epsilon - \epsilon_\theta(x_t, t)\|^2 \right] \tag{2.3}$$

where $\mathbb{E}$ denotes the expectation over the distribution of original data x_0, the noise ϵ, and the time step t.

The optimization of this loss function allows the diffusion model to learn an accurate reverse process, effectively denoising the data and reconstructing the original signal from its noisy version. By iterating through this reverse process, the model can generate new samples that are consistent with the learned data distribution, showcasing the powerful generative capabilities of diffusion models.

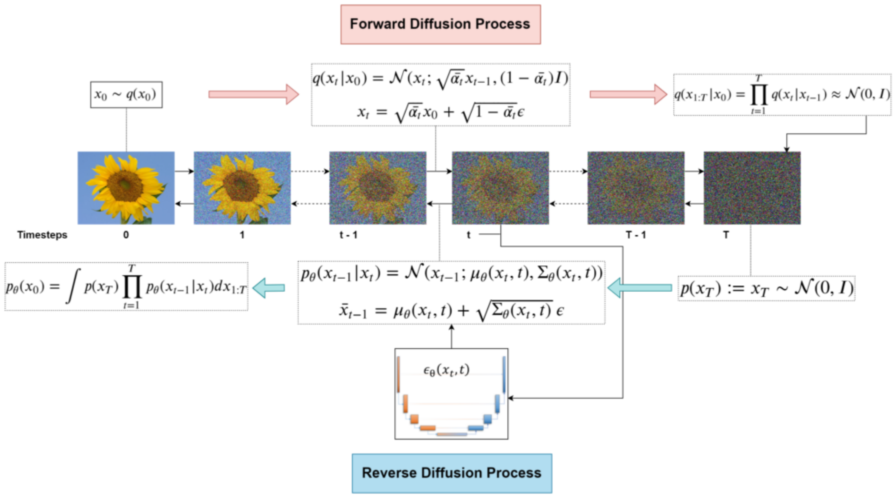

Figure 2.3: The mathematical framework depicting the forward and reverse diffusion processes. Adapted from **OpenCVDDPM**.

2.1.2 Applications of Diffusion Models

Diffusion models have found applications in a wide array of fields beyond just medical imaging. This versatility underscores their broad applicability and effectiveness in generating high-quality synthetic data. Below, we explore some prominent areas where diffusion models have been successfully applied.

Natural Image Generation

Diffusion models have made significant strides in the field of natural image generation. They can create detailed and diverse images that closely mimic the distribution of real-world photographs. This capability has implications for content creation, gaming, and virtual reality, providing a way to generate landscapes, objects, and characters that are indistinguishable from real ones.

$$x_{synth} = \text{DiffusionModel}(z), \quad z \sim \mathcal{N}(0, I) \tag{2.4}$$

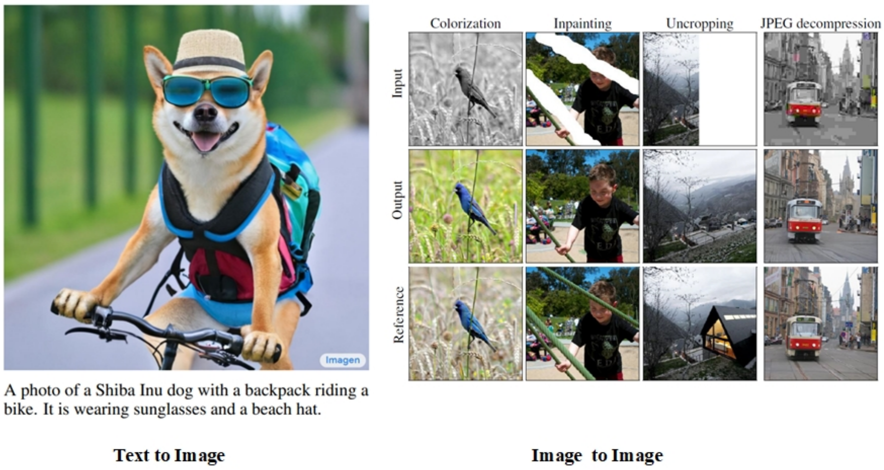

FIGURE 2.4: Examples of some applications for diffusion models in different domains (Saharia et al. 2022).

Here, x_{synth} represents the synthetic image generated by the diffusion model, and z is a sample from a standard Gaussian distribution used as input noise.

Text-to-Image Synthesis

Another groundbreaking application of diffusion models is in text-to-image synthesis, where models generate images directly from textual descriptions. This application merges natural language processing with image synthesis, opening new avenues in automated content creation and aiding in tasks such as storyboarding and concept art creation.

$$x_{image} = \text{DiffusionModel}(text\_description) \tag{2.5}$$

In this context, x_{image} is the image generated from a given text description through the diffusion process.

Audio Synthesis

Diffusion models are also revolutionizing the field of audio synthesis, including voice generation and music production. They can generate realistic and coherent audio clips from a variety of inputs or even from noise, contributing to applications such as virtual assistants, speech synthesis for individuals with speech impairments, and new music creation.

$$x_{audio} = \text{DiffusionModel}(audio\_input) \tag{2.6}$$

Here, x_{audio} denotes the synthesized audio generated by the model, showcasing the potential of diffusion models in creating diverse soundscapes and spoken content.

Molecular Design

In the realm of chemistry and drug discovery, diffusion models are employed for the generation and optimization of molecular structures(Yim et al. 2024). By learning the distribution of viable molecular configurations, these models can propose new compounds with desired properties, speeding up the drug development process and contributing to personalized medicine.

$$x_{molecule} = \text{DiffusionModel}(property\_constraints) \tag{2.7}$$

$x_{molecule}$ represents the molecular structure generated to meet specific property constraints, demonstrating the model's application in scientific research and pharmaceuticals.

2.1.3 Key Advancements and Milestones in Diffusion Models

The development of diffusion models has been marked by numerous significant advancements and milestones. In this section, we chronicle the pivotal breakthroughs and notable contributions that have shaped the landscape of diffusion models.

Initial Developments

The initial concept of diffusion models was inspired by non-equilibrium thermodynamics and the stochastic diffusion process. Early works laid the groundwork by exploring how random noise can be systematically added and then removed from data to learn intricate data distributions (Sohl-Dickstein et al. 2015).

Introduction of Score-Based Models

A significant leap forward was the development of score-based generative models by (Yang et al. 2023), which utilized gradients (scores) of the data distribution to guide the diffusion process, enhancing the model's ability to generate high-quality synthetic data.

Development of Denoising Diffusion Probabilistic Models

The introduction of Denoising Diffusion Probabilistic Models (DDPMs) by (Ho et al. 2020). marked another milestone, presenting a robust framework that combined variational inference with diffusion processes, significantly improving the quality and efficiency of generated samples.

$$x_0 = \mathrm{DDPM}(x_t, \epsilon_t; \theta), \quad \epsilon_t \sim \mathcal{N}(0, I) \tag{2.8}$$

θ denotes the parameters of the DDPM, illustrating the model's denoising capability.

Improvements in Sampling Efficiency

Further improvements in diffusion models aimed at increasing sampling efficiency, such as Fast Sampling Algorithms introduced by Song et al., which reduced the number of required sampling steps without compromising output quality, marking a crucial advancement for practical applications.

Extension to Diverse Data Types

Advancements have also been seen in extending diffusion models beyond images to other data types, including audio, text, and molecular structures, demonstrating the model's versatility and adaptability to various domains.

Integration with Other AI Techniques

Recent breakthroughs include integrating diffusion models with other AI techniques, such as reinforcement learning and GANs, to create hybrid models that leverage the strengths of each approach, pushing the boundaries of generative modeling.

Latest Innovations and Ongoing Research

The field continues to evolve rapidly, with ongoing research exploring novel architectures, optimization techniques, and applications of diffusion models. These developments are setting new benchmarks for what can be achieved with generative modeling.

Each of these milestones has contributed to the maturation of diffusion models, turning them into one of the most promising areas in generative modeling. By understanding these key advancements, we can appreciate the rapid evolution of diffusion models and their growing impact on various fields, including medical imaging.

2.1.4 Challenges and Limitations of Diffusion Models

Despite the significant advancements in diffusion models, they are not without their challenges and limitations. This section explores the various hurdles currently faced by practitioners and researchers in the field, setting realistic expectations for the technology's capabilities and applications.

Computational Intensity

One of the major challenges associated with diffusion models is their computational intensity. The iterative nature of the reverse diffusion process requires substantial computational resources, particularly in terms of memory and processing power.

$$T_{total} = \sum_{t=1}^{N} T_{step}(t) \tag{2.9}$$

Here, T_{total} represents the total computation time, N is the number of steps, and $T_{step}(t)$ denotes the time taken for each step.

Long Training Duration

Diffusion models typically require extended periods of training to achieve satisfactory results. This is particularly challenging when dealing with large datasets or aiming to generate high-resolution outputs.

Quality of Generated Samples

While diffusion models can generate high-quality samples, maintaining consistency and avoiding artifacts remain significant challenges. The balance between noise reduction and detail preservation is delicate, often resulting in trade-offs.

Model Generalization

Generalizing diffusion models to diverse datasets and different types of noise is a persistent challenge. Models trained on specific types of data may not perform well when exposed to new or unseen data types or noise distributions.

Application-Specific Challenges

Diffusion models face unique challenges in specific applications, such as medical imaging or audio synthesis, where high fidelity and accuracy are crucial. Tailoring diffusion models to meet the stringent requirements of these applications requires extensive customization and fine-tuning.

Ethical and Privacy Concerns

The generation of realistic data samples, especially in domains like medical imaging, raises ethical and privacy concerns. Ensuring that the synthetic data generated by diffusion models do not infringe on privacy rights or ethical standards is an ongoing challenge.

Scalability

Scaling diffusion models to accommodate larger datasets or higher-dimensional data without compromising performance or increasing computational demands poses another significant hurdle.

Each of these challenges represents a barrier to the widespread adoption and application of diffusion models. Addressing these limitations requires concerted efforts from the research community, ongoing innovation, and the development of more efficient and adaptable models. By acknowledging and tackling these issues, we can further unlock the potential of diffusion models and extend their applicability across a broader range of fields and applications.

2.1.5 Summary

This chapter has provided an overview of diffusion models, covering their introduction, mathematical framework, diverse applications, key advancements, and the challenges they face. We have explored how these models have evolved from theoretical constructs into powerful tools for generative tasks across different domains, including image and audio synthesis.

Chapter 3

PALETTE: A New Frontier in Image-to-Image Diffusion Models

3.1 Introduction

The landscape of generative models has undergone remarkable transformation recently, especially with the advent of diffusion models that have significantly impacted the fields of computer vision and image processing. This evolution is well captured in the study by Saharia, Chitwan, et al., titled "Palette: Image-to-image diffusion models," published in the ACM SIGGRAPH 2022 conference proceedings. Here, diffusion models are spotlighted as robust contenders to Generative Adversarial Networks (GANs), offering superior performance across a range of applications from image synthesis to advanced super-resolution techniques.

PALETTE emerges from this innovative lineage as an exemplary implementation of image-to-image diffusion models. It is distinguished by its unmatched flexibility and efficiency, designed to accommodate an extensive array of translation tasks seamlessly. Unlike traditional models, PALETTE's core methodology is predicated on reverse diffusion, which systematically evolves a sample from a simplistic Gaussian distribution into intricate, detailed data representations. This is achieved through iterative denoising phases, meticulously orchestrated by neural networks, bestowing upon PALETTE commendable stability during training phases and the ability to embrace a broad spectrum of data distribution patterns. This unique attribute significantly diminishes the common pitfalls such as mode collapse, which frequently afflict GANs. Below is an illustrative comparison showcasing PALETTE's effectiveness alongside other colorization methods.

FIGURE 3.1: Illustration of colorization methods on ImageNet validation images. Comparisons include PixColor, ColTran, a regression baseline, PALETTE (our approach), and the original reference images. Image adapted from "Palette: Image-to-image diffusion models" by Saharia, Chitwan, et al., ACM SIGGRAPH 2022.

As demonstrated, PALETTE excels in producing images that closely match the reference, significantly enhancing realism and detail compared to other colorization techniques. PALETTE's prowess is not confined to generic image-generation tasks. It extends its utility to specific applications critical in real-world scenarios, including colorization, inpainting, uncropping, and JPEG artifact removal, showcasing its universal appeal. A particularly noteworthy application, and the focus of this chapter, is leveraging PALETTE's sophisticated colorization capabilities to enhance the quality of low-dose CT images. The hallmark features that propel PALETTE to the forefront of diffusion models are multifold:

- **Task Agnosticism:** PALETTE introduces a paradigm of universal application across diverse image-to-image translation tasks, thus obviating the requirement for specialized tuning and architecture adjustments traditionally necessary in other models.

- **Training Stability and Efficiency:** The framework of PALETTE is fortified with innovative training methodologies and loss functions, fostering an environment of remarkable stability and efficiency during training sessions. This contrasts sharply with the challenges of noise overfitting and training divergence prevalent in other diffusion models.

- **High-Quality Image Synthesis:** Through the meticulous optimization of the diffusion process, PALETTE consistently generates images of unparalleled quality, showcasing enhanced details and precise colorization, thereby establishing new benchmarks in the fidelity of image-to-image translations.

- **Mode Diversity:** PALETTE's algorithm is adept at capturing the extensive variability inherent in image distributions, ensuring the production of outputs that are as diverse as they are realistic, and overcoming limitations encountered in previous generative models.

- **Computational Efficiency:** Despite the naturally iterative essence of diffusion processes, PALETTE introduces significant improvements in computational efficiency, thereby democratizing access to high-quality image translation for broader applications.

This chapter sets the stage for an exploration of how PALETTE, particularly through its colorization functionality, can enhance low-dose CT images.

3.1.1 Model Architecture

The PALETTE model employs a unique architecture tailored for image-to-image diffusion processes, primarily focusing on tasks like colorization, inpainting, uncropping, and JPEG artifact removal. Its architecture is designed to accommodate a wide range of image-to-image translation tasks without task-specific modifications. Specifically, PALETTE utilizes a modified U-Net architecture, a deviation from the traditional 256 * 256 class-conditional U-Net model. The key modifications include:

- **Absence of class-conditioning:** Unlike conventional U-Net models that might utilize class information to guide the image generation process, PALETTE operates without class-conditional inputs, enhancing its versatility across various tasks.

- **Source image conditioning:** PALETTE incorporates source image conditioning through concatenation, enabling the model to adapt more effectively to different image-to-image translation scenarios.

Furthermore, PALETTE integrates self-attention mechanisms, an essential component in enhancing the model's performance for complex image-to-image translation tasks. This inclusion is based on empirical studies showing that self-attention layers significantly contribute to the overall quality and diversity of the generated samples.

3.1.2 Loss Functions

PALETTE employs a multifaceted approach to loss functions, tailored to the intricate demands of image-to-image translation tasks. The model's strategy encompasses a blend of different loss types to strike an optimal balance between fidelity to the original image and diversity in the generated results.

Firstly, the L1 loss, or mean absolute error, is leveraged for its effectiveness in promoting pixel-wise accuracy, which directly contributes to the fidelity of the translated images:

$$\mathcal{L}_{L1}(\theta) = \mathbb{E}_{y_0,y_t,t}\left[\|f_\theta(y_t,t) - y_0\|_1\right]. \tag{3.1}$$

The L1 loss ensures that the generated images maintain structural and content integrity relative to the target images by minimizing the absolute differences at the pixel level.

Additionally, PALETTE incorporates the L2 loss, or mean squared error, which penalizes larger discrepancies between the generated and target images more severely than smaller ones, contributing to smoother gradients during training:

$$\mathcal{L}_{L2}(\theta) = \mathbb{E}_{y_0,y_t,t}\left[\|f_\theta(y_t,t) - y_0\|_2^2\right]. \tag{3.2}$$

This component of the loss function aids in reducing variance and ensuring overall coherence in the generated images.

Beyond L1 and L2 losses, PALETTE also explores the use of adversarial losses to introduce a competitive dynamic between the generative model and a discriminator model. The adversarial loss encourages the generation of images that are not only close to the original in a pixel-wise sense but are also indistinguishable from real images in the distribution space:

$$\mathcal{L}_{adv}(\theta, \phi) = \mathbb{E}_{y_0, y_t}\left[\log D_\phi(y_0) + \log(1 - D_\phi(f_\theta(y_t, t)))\right], \tag{3.3}$$

where D_ϕ represents the discriminator, parameterized by ϕ, and f_θ is the generative model, parameterized by θ. This adversarial component fosters innovation in generated samples, enhancing their naturalness and diversity. Beyond L1, L2, and adversarial losses, the model incorporates a noise prediction loss to enhance its capability in stochastic modeling. The noise prediction loss ensures that the model accurately predicts the noise added during each step of a generative process, critical for models like diffusion models where reversing the noise addition is essential:

$$\mathcal{L}_{noise}(\theta) = \mathbb{E}[(\epsilon_{\text{predicted}} - \epsilon_{\text{actual}})^2], \tag{3.4}$$

where $\epsilon_{\text{predicted}}$ represents the noise predicted by the model, parameterized by θ, and ϵ_{actual} is the actual noise applied to the data during training. This loss term helps in fine-tuning the model's ability to revert the noising process accurately, improving its generative performance and fidelity to the original data distribution. The integration of this component is crucial for achieving high-quality synthesis and regeneration capabilities in models dealing with noisy or complex data structures.

Combining these elements, PALETTE's total loss function amalgamates the advantages of each loss component, formulated as:

$$\mathcal{L}_{total}(\theta, \phi) = \lambda_{L1}\mathcal{L}_{L1}(\theta) + \lambda_{L2}\mathcal{L}_{L2}(\theta) + \lambda_{adv}\mathcal{L}_{adv}(\theta, \phi) + \lambda_{noise}\mathcal{L}_{noise}(\theta), \tag{3.5}$$

where λ_{L1}, λ_{L2}, λ_{adv}, and λ_{noise} are weights assigned to each respective loss term, allowing for fine-tuning of their influence on the overall training process.

By judiciously combining these loss functions, PALETTE achieves a delicate balance between adhering closely to the source images and introducing diversity into the generated images, thereby addressing the dual challenges of fidelity and variation inherent in image-to-image translation tasks.

3.2 Application of Palette to Low-Dose to High-Dose CT Image Enhancement

In this section, we discuss the application of Palette for enhancing low-dose CT images to achieve high-dose CT image quality. This process is vital for reducing radiation exposure while maintaining or even improving the diagnostic quality of the images.

3.2.1 Model Architecture Adaptation

The Palette model, initially designed for various image-to-image translation tasks, has been adapted to tackle the low-dose to high-dose CT image enhancement challenge. This adaptation involves fine-tuning the standard Palette architecture to better handle the specific noise and detail characteristics inherent in low-dose CT images.

3.2.2 Loss Function Modification

To address the low-dose to high-dose CT translation, Palette utilizes an adapted set of loss functions. While maintaining the standard L1 and L2 losses for baseline fidelity, additional loss components such as $\mathcal{L}_{adv}$, and $\mathcal{L}_{noise}$ be introduced to target the specific types of noise and artifacts characteristic of low-dose CT scans. The balance between these loss components is critical for achieving an optimal enhancement effect.

3.2.3 Training Protocol for CT Image Enhancement

The training protocol for this particular application of Palette follows the general diffusion model training practices but is specifically augmented to cater to CT images' unique characteristics. This includes selecting an appropriate dataset consisting of paired low-dose and high-dose CT images and employing specialized data augmentation techniques relevant to medical imaging.

3.3 Results and Analysis

Following the application of Palette to enhance low-dose to high-dose CT imaging, we evaluate the model's performance using established quantitative metrics: Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM). These metrics serve as objective indicators of the quality of enhancement and effectiveness in noise reduction.

3.3.1 Quantitative Results

Below, we present a comparison between the original high-dose CT images and their corresponding images enhanced by Palette. The table showcases the improvement in image quality as quantified by the PSNR and SSIM values:

Table 3.1: Quantitative Palette-Enhanced Images

Metric	Palette-Enhanced High-Dose
PSNR (dB)	24.5382
SSIM	0.48

The PSNR and SSIM values were calculated based on the comparison between the original high-dose images and their Palette-enhanced counterparts. Higher values indicate better image quality, with PSNR measuring the peak error, and SSIM evaluating the similarity in structural information between the two images.

3.3.2 Qualitative Evaluation

In addition to quantitative metrics, we conduct a qualitative evaluation, assessing the visual improvements in the enhanced images. This involves comparing the clarity, detail visibility, and noise levels in low-dose original and Palette-enhanced images. This section showcases the visual outcomes of applying Palette to enhance low-dose CT images. The figures below display a comparison between original low-dose CT images and their enhanced counterparts after being processed by Palette. These figures illustrate the effectiveness of Palette in enhancing the clarity

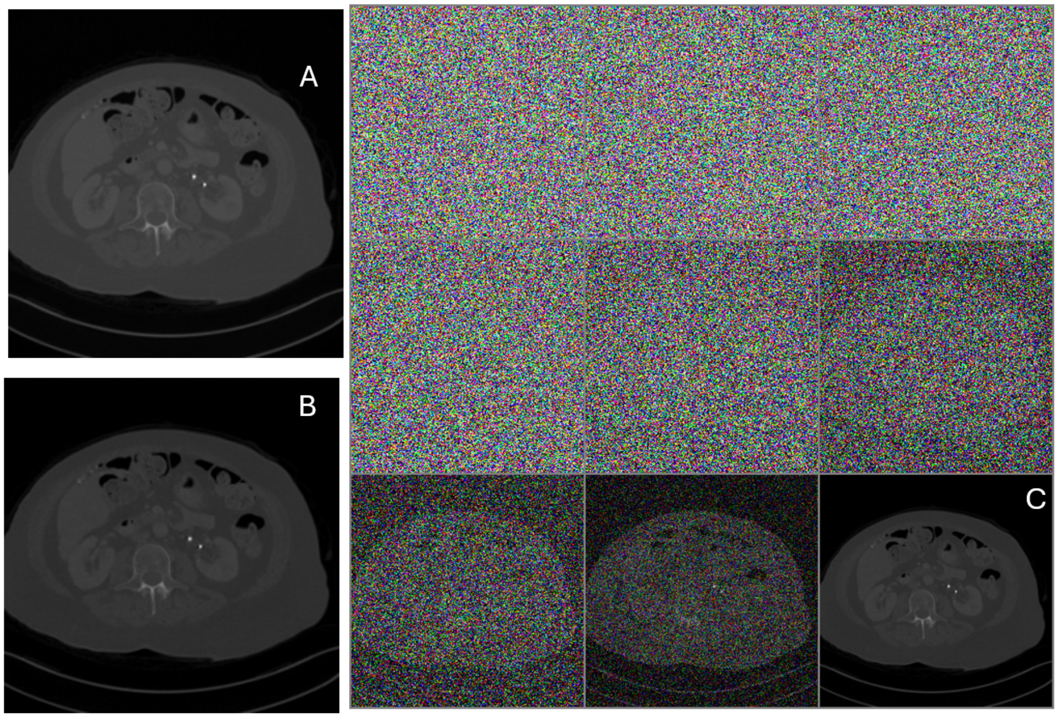

Figure 3.2: Comparative visualization of the CT image enhancement using PALETTE: (A) Original high-dose, (B) Enhanced by PALETTE, and (C) Visualization of the diffusion process

and detail of CT images while reducing noise, making them more suitable for diagnostic purposes. The visual comparison highlights the improvements in image quality, demonstrating Palette's potential in aiding the medical field, particularly in scenarios where reducing patient exposure to radiation is crucial.

3.4 Conclusion

This exploration of Palette's application to enhancing low-dose CT images has demonstrated its potential to improve image quality while mitigating radiation exposure risks. The findings from this study suggest avenues for further refinement and validation, with the ultimate aim of integrating such advanced image processing techniques into practical applications for safer and more accurate medical imaging.

Chapter 4

Noise-Characterized Diffusion Model

This chapter introduces a mathematical formulation designed to enhance the quality of low-dose computed tomography (CT) images. By applying diffusion models integrated with explicit noise characterization, we adapt and extend traditional diffusion model equations to tackle the unique noise patterns observed in low-dose CT scans.

4.1 Forward Process with Noise Characterization

The forward process in noise-characterized diffusion models plays a pivotal role in enhancing the quality of low-dose CT images. Unlike traditional approaches, our method introduces noise into clean images systematically over T iterations.

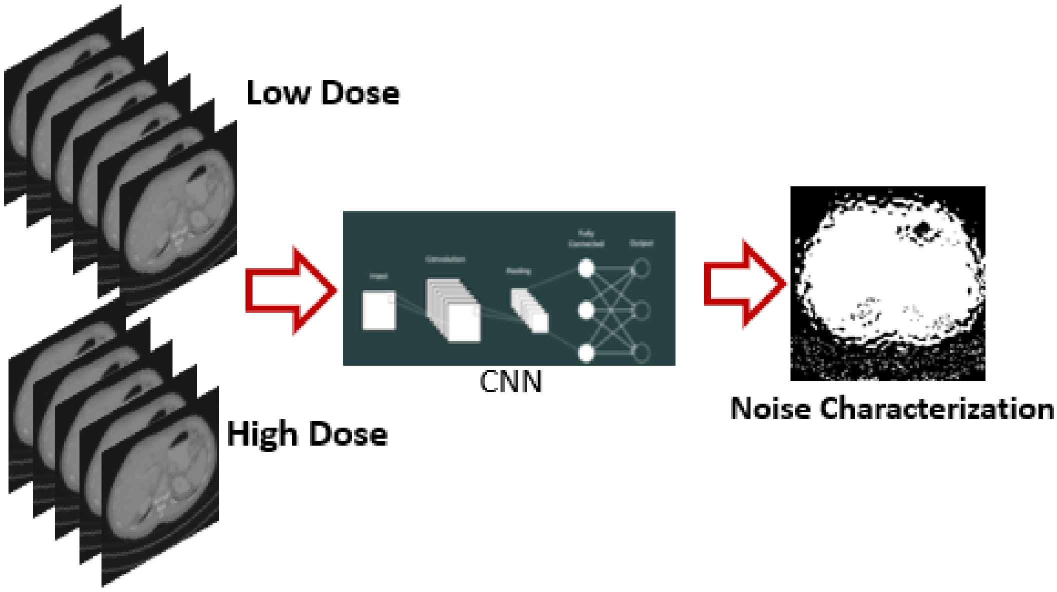

FIGURE 4.1: Noise Characterization using CNN

This section delves into the methodology, mathematical foundation, and practical implications of our novel forward process with noise characterization.

4.1.1 Conceptual Framework

The conceptual framework behind the forward process is rooted in the understanding of noise dynamics within low-dose CT imaging. Low-dose CT scans are inherently noisier compared to their high-dose counterparts due to reduced X-ray photon counts, leading to increased statistical noise. This noise follows a specific distribution characteristic of the imaging system and the radiation dose level. Recognizing tahis, we developed a noise model, denoted as N_c, that aims to replicate this specific noise distribution accurately. In our approach, the clean image is transformed into T iterations, where T represents the total number of steps in the forward diffusion process. At each step, noise is incrementally added to the image,

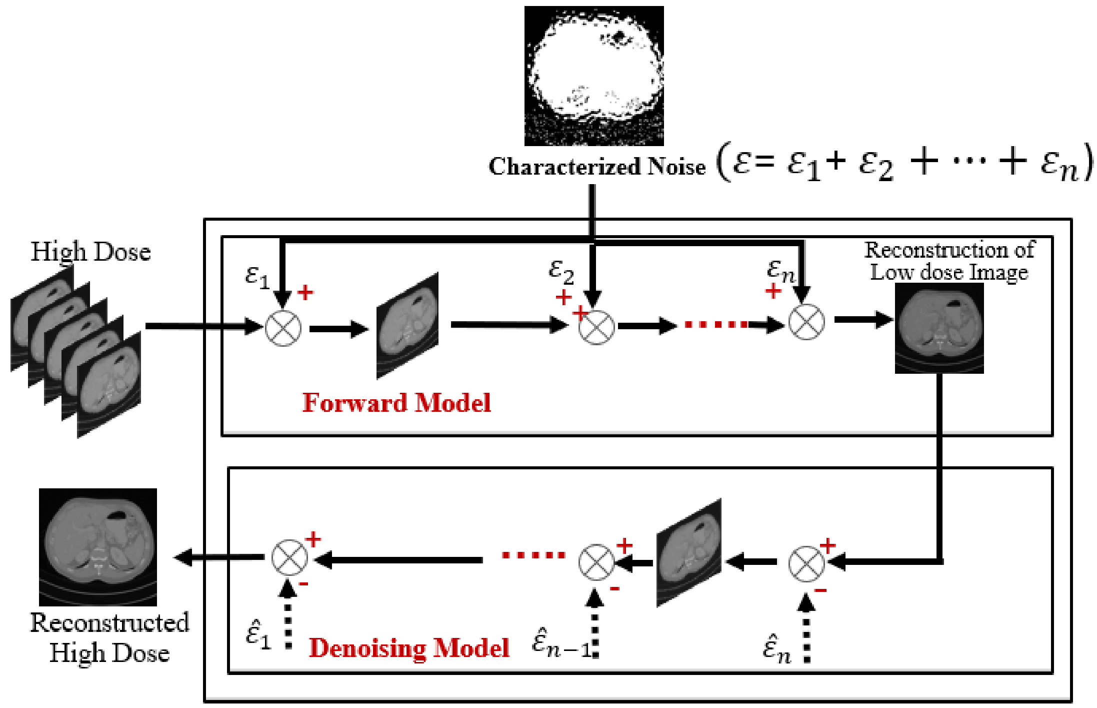

FIGURE 4.2: Proposed architecture for enhancing quality of low-dose CT scans via generative diffusion model

simulating the increase in noise levels typically observed in low-dose CT scans. The objective is to reach a noise level consistent with actual low-dose CT images while maintaining control over the noise characteristics introduced at each stage.

4.1.2 Mathematical Representation

The mathematical representation of our forward process is encapsulated in the following equation:

$$q(y_{t+1}|y_t) = N_c(y_t; \alpha_t y_t, (1 - \alpha_t)C_t), \tag{4.1}$$

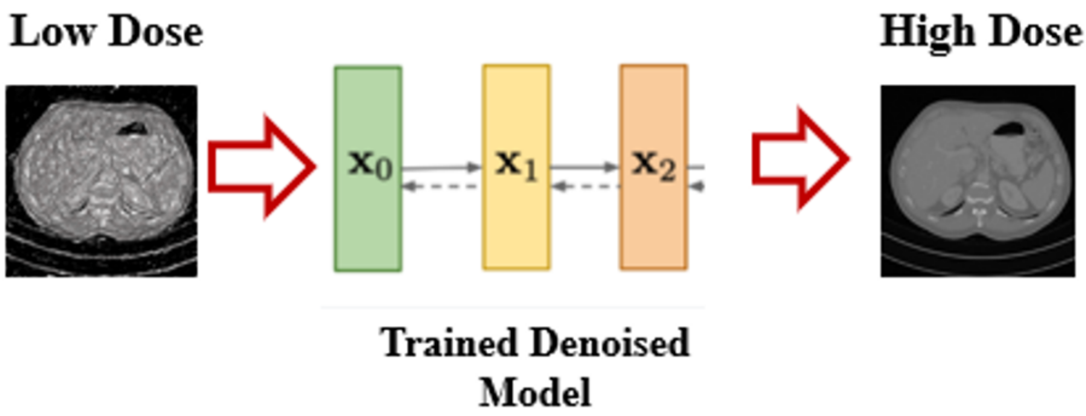

Figure 4.3: Proposed architecture for enhancing quality of low-dose CT scans via generative diffusion model

where y_t represents the image after t iterations, and y_{t+1} is the image at the subsequent iteration. The function N_c represents the noise model applied to the image, governed by parameters α_t and C_t. Here, α_t modulates the degree of the original signal retained in each step, while $(1 - \alpha_t)C_t$ determines the variance of the noise introduced.

The parameter α_t is a critical component of our model as it guides the progression of noise addition. Initially set to a value close to one, α_t decreases with each iteration, allowing for a gradual increase in noise. The choice of α_t is based on empirical data and simulation studies designed to mimic the noise escalation in low-dose CT scans.

The covariance matrix C_t is equally significant as it characterizes the noise distribution's nature. In the context of CT imaging, noise can exhibit varying behaviors, such as being signal-dependent or possessing spatial correlations. By incorporating C_t, our model acknowledges and replicates these complexities, enabling a more accurate simulation of low-dose CT noise.

The iteration over T steps necessitates a balance between computational efficiency and noise simulation fidelity. Too few iterations may lead to underdeveloped noise patterns, while too many could result in computational inefficiency or excessive noise. Therefore, determining the optimal value of T is subject to a trade-off between computational resources and model accuracy.

4.1.3 Implications for Low-Dose CT Imaging

The implications of accurately modeling the forward process in low-dose CT imaging are profound. By precisely characterizing and replicating the noise inherent in low-dose scans, our approach enables a deeper understanding of noise effects on image quality and diagnostic accuracy. Furthermore, this model serves as a foundation for the subsequent reverse process, where the aim is to recover the clean image from its noisy counterpart.

In summary, the forward process with noise characterization is a cornerstone of our approach to enhancing low-dose CT images. Through careful mathematical modeling and practical implementation, we can simulate the noise characteristics of low-dose CT scans accurately, setting the stage for effective noise reduction and image enhancement in subsequent steps of the diffusion model framework

4.2 Reverse Process with Noise Characterization

In the realm of enhancing low-dose Computed Tomography (CT) images through diffusion models, the reverse process stands as a critical component. In the reverse process, the primary goal is to reconstruct the original clean image from its noisy version. This is achieved by employing a specialized neural network f_θ, meticulously designed to counteract the specific noise attributes recognized during the forward process.

4.2.1 Conceptual Foundation

The reverse process is conceptually the antithesis of the forward diffusion phase. While the forward phase incrementally introduces noise into the clean images, the reverse phase systematically eliminates this noise, aiming to restore the original quality of the images. This denoising procedure is pivotal, as it directly impacts the final image quality and, consequently, the diagnostic value of the CT scans.

The neural network f_θ is at the heart of this process. It is not just a generic denoising tool but a sophisticated algorithm tailored to the unique noise characteristics identified in the low-dose CT images. This customization allows f_θ to target and mitigate specific noise statistics, a capability beyond that of standard denoising techniques.

4.2.2 Mathematical Formulation

The effectiveness of the reverse process hinges on the proper formulation of the learning objective, represented as:

$$L(\theta) = \mathbb{E}_{(x,y),(\epsilon,y)} \left[||f_\theta(x, \gamma y_0 + (1-\gamma)\epsilon, \gamma) - \epsilon|| \right], \tag{4.2}$$

where x denotes the original clean image, y_0 the observed noisy image, and ϵ the noise element. The variable γ serves as a scaling factor, adjusting the influence of the original and noisy components during training. This loss function is engineered to refine f_θ's capability to discern and reverse the noise pattern imprinted during the forward phase.

4.2.3 Implementation Challenges

Implementing the reverse process entails several challenges, notably the accurate training of the neural network f_θ. The network must learn to distinguish between noise-induced artifacts and intrinsic image details, a task of considerable complexity given the subtle nature of certain pathological markers.

Another challenge lies in the selection of the parameter γ, which must be finely tuned to balance the noise and signal components effectively. This balance is crucial for preserving essential image features while removing noise, a delicate equilibrium that is vital for maintaining diagnostic accuracy.

4.3 Training the Model

Training the noise-characterized denoising model, f_θ, is a crucial step towards achieving enhanced image quality in low-dose Computed Tomography (CT) imaging. This section delves into training methodology that leverages a characterized noise model to simulate realistic noisy observations. Our objective is to refine f_θ to proficiently reverse the noise process, thereby restoring the high-quality characteristics of the original image from its degraded version.

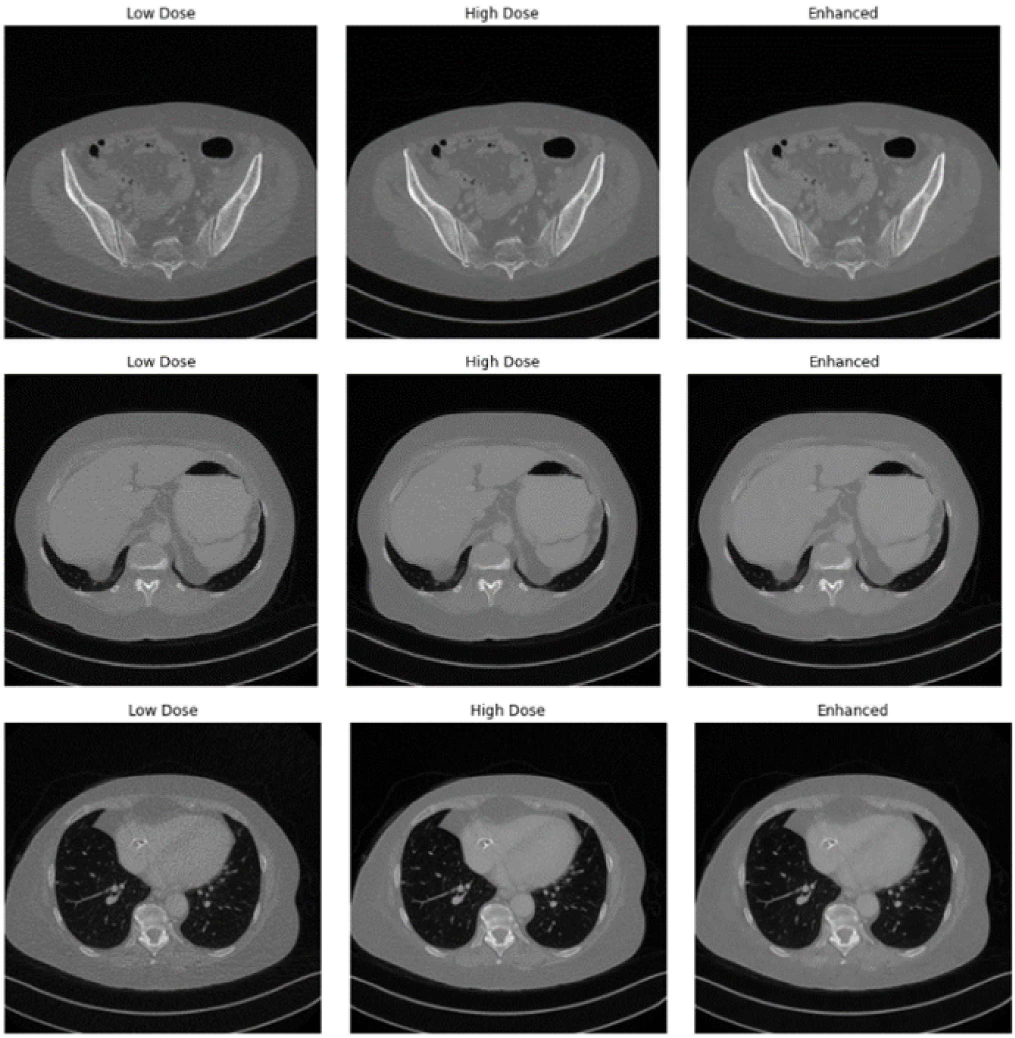

FIGURE 4.4: Lefts: LDCT test image | Rights: Generated HDCT Test Image

4.3.1 Training Objective

The training procedure revolves around the optimization of the loss function $L(\theta)$, which is designed to minimize the discrepancy between the denoised image and the original clean image. The model's performance hinges on the effective reduction of this loss, reflecting the denoising capabilities of f_θ. This optimization is performed across a diversified dataset comprising pairs of low- and high-dose CT images, facilitating a comprehensive learning environment that encompasses a wide range of noise patterns and imaging scenarios.

4.3.2 Algorithmic Approach

The training algorithm employs a structured iterative process to fine-tune the parameters of f_θ. The sequence of operations is encapsulated in the following algorithm 1.

Algorithm 1 Training a noise-characterized denoising model f_θ

1: **repeat**
2: $(x, y_0) \sim p(x, y)$ ▷ Sample from the data distribution
3: $\gamma \sim p(\gamma)$ ▷ Sample noise level
4: $\epsilon \sim N_c(\epsilon; C_t)$ ▷ Sample noise from the characterized model
5: $y_t = \sqrt{\gamma} y_0 + \sqrt{1-\gamma}\epsilon$ ▷ Apply noise to the clean image
6: Take a gradient descent step on:
7: $\nabla_\theta \| f_\theta(x, \sqrt{\gamma} y_0 + \sqrt{1-\gamma}\epsilon, \gamma) - \epsilon \|$
8: **until** convergence

This algorithm underscores the cyclical nature of the training process, iteratively updating f_θ to better model the inverse of the noise application process. By alternating between sampling from the characterized noise distribution and updating the model parameters, we systematically drive the model towards optimal noise reduction performance.

4.3.3 Challenges in Training

Training a denoising model, particularly one characterized by low-dose CT noise, presents unique challenges. These include ensuring a diverse and representative training set, managing overfitting, and effectively simulating realistic noise conditions.

4.3.4 Evaluation Metrics

Throughout the training process, the performance of f_θ is meticulously evaluated using a suite of metrics, including but not limited to, the Peak Signal-to-Noise Ratio (PSNR) and the Structural Similarity Index Measure (SSIM).

The training iteratively refines the model parameters to adapt to the characterized noise features, thus preparing the model for effective denoising of low-dose CT images.

4.4 Iterative Refinement

In the inference phase, our methodology employs the reverse process iteratively to refine the noisy images progressively towards a high-fidelity reconstruction. This iterative refinement process is crucial as it ensures that the denoising closely aligns with the actual noise properties characteristic of low-dose Computed Tomography (CT) scans.

The core of this iterative refinement lies in employing the learned model, f_θ, to gradually eliminate the noise from the low-dose CT images. This process is informed by the characterized noise at each iteration, enabling the model to effectively reverse the diffusion process introduced during the forward phase.

The refinement procedure is detailed as follows:

1. Start with an initial estimate of the noisy image, which, in the context of this process, is the output of the forward diffusion step applied to the clean, high-dose image.

2. At each iteration t, apply the learned denoising model f_θ to the current estimate of the noisy image to predict the noise component $\hat{\epsilon}_t$.

3. Update the image estimate by removing the predicted noise from the current estimate: $y_{t-1} = f_\theta^{-1}(y_t, \hat{\epsilon}_t, \gamma_t)$, where f_θ^{-1} represents the inverse of the forward model applied by f_θ.

4. Repeat steps 2 and 3 for a fixed number of iterations or until a convergence criterion is met, such as a minimal change between consecutive image estimates or reaching a pre-defined level of noise reduction as measured by appropriate metrics such as PSNR (Peak Signal-to-Noise Ratio) or SSIM (Structural Similarity Index Measure).

This iterative process is designed to refine the image progressively, ensuring each step reduces the noise level while preserving the critical details and structural integrity of the original CT images. The refinement continues until the reconstructed image closely approximates the quality and fidelity of a high-dose CT scan, thereby achieving the objective of enhancing the diagnostic quality of low-dose CT images without additional radiation exposure to the patient.

The effectiveness of this iterative refinement process is measured by comparing the denoised image with the original high-dose CT image, evaluating both qualitative aspects, such as visual clarity and detail preservation, and quantitative metrics, including PSNR and SSIM.

4.5 Conclusion

The development and implementation of the proposed noise-characterized diffusion model (NCDM) represent a new approach to the enhancement of low-dose Computed Tomography (CT) imaging. The cornerstone of this model is its utilization of detailed noise characterization combined with advanced diffusion processes. By specifically targeting the unique noise patterns found in low-dose CT scans, the NCDM provides a mechanism to significantly improve image quality, fidelity, and diagnostic value without increasing radiation dosage. Key to the success of this model is its ability to adapt and refine traditional diffusion equations to suit the specific challenges presented by CT imaging noise. The incorporation of a forward and reverse process, informed by precise noise characteristics, enables the effective simulation and removal of noise while preserving essential anatomical details.

In conclusion, the noise-characterized diffusion model is an enhancement in the domain of medical diagnostics and patient safety.

Chapter 5

Comparative Analysis of Low-Dose CT-Image Enhancement Techniques

In this chapter,new objective is to conduct a comparative analysis that evaluates the efficacy of PALETTE and NCDM vis-à-vis CYCLE-GANs. The emphasis on CYCLE-GANs, inspired by the pioneering work of McCollough et al. (McCollough et al. 2020) and further discussed in the MATLAB documentation (MathWorks 2023), sets a high benchmark for performance in enhancing the quality of low-dose CT (LDCT) images. By comparing these approaches, we aim to uncover insights that could lead to improvements in diagnostic imaging, ensuring both the safety of patients through minimized radiation exposure and the reliability of medical diagnoses.

5.1 Low-Dose CT-Image Enhancement Using CYCLE-GAN

This section explores the use of a CYCLE-GAN neural network for the enhancement of noisy low-dose CT images to high-quality high-dose CT images. The approach leverages low-dose CT images as the source domain and regular-dose CT images as the target domain (MathWorks 2023). The CYCLE-GAN consists of two generators, G_{XY} and G_{YX}, along with two discriminators, D_X and D_Y. Generator G_{XY} aims to transform low-dose CT images into high-dose CT images, whereas G_{YX} attempts to reconstruct the original low-dose images from the enhanced high-dose images. Discriminators D_X and D_Y assess the authenticity of the low-dose and high-dose images, respectively.

The process begins with an original low-dose CT image, which is fed into G_{XY}, resulting in a synthesized high-dose image. This image is then used by G_{YX} to restore the low-dose CT image, ensuring cycle consistency. The high-dose CT image, provided for reference, enables quality assessment of the enhanced images. This cyclic transformation is crucial for maintaining the essential content within the images, which is vital for accurate medical diagnosis (MathWorks 2023).

In the study conducted by (You et al. 2018), a structurally sensitive multi-scale deep neural network was utilized for low-dose CT denoising, exhibiting significant improvements in image quality. The network achieved notable PSNR and SSIM values, with a PSNR of 29.679 dB and an SSIM of 0.813, indicating the effectiveness of the approach in producing high-quality CT images (You et al. 2018).

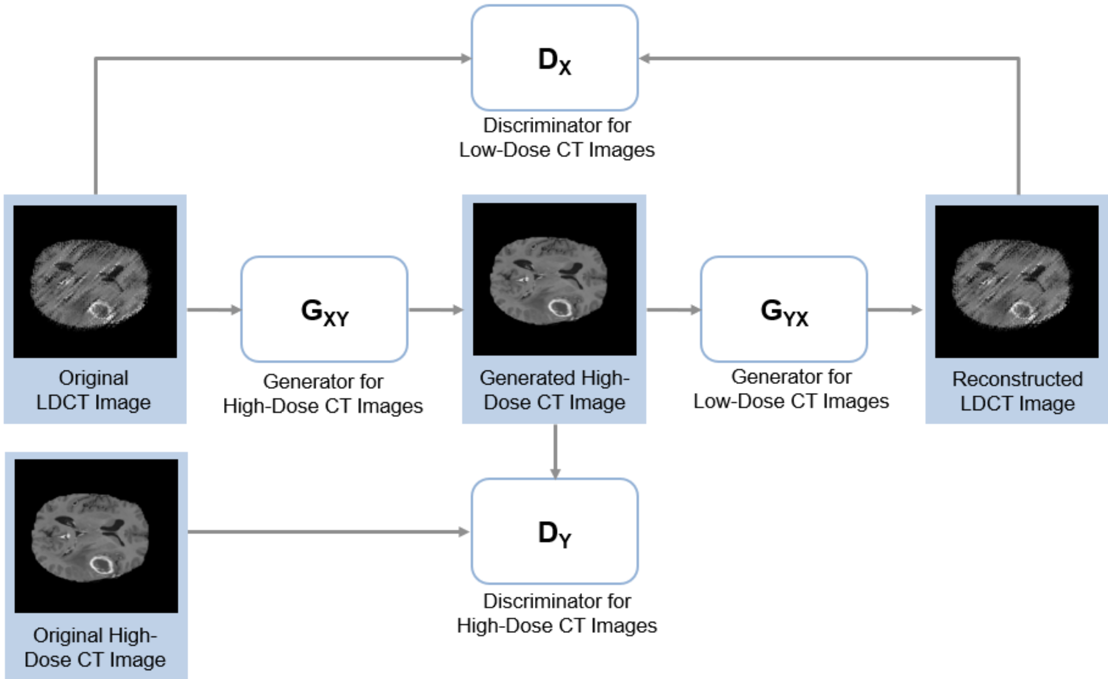

FIGURE 5.1: The architecture of a CYCLE-GAN neural network applied to CT image enhancement.(MathWorks 2023)

5.2 Comparison of PSNR and SSIM Values

In this section, we evaluate the performance of our Noise-Characterized Diffusion Model (NCDM) against CYCLE-GAN and Palette for enhancing low-dose CT images. The evaluation metrics used for comparison are the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM), which are standard metrics for measuring image quality and similarity, respectively.

5.2.1 PSNR Comparison

PSNR, or Peak Signal-to-Noise Ratio, is a widely utilized metric for assessing the quality of reconstruction in lossy compression algorithms. It quantifies the relationship between the highest possible signal power and the power of the corrupting noise that compromises the accuracy of the signal's representation. Table 5-1 presents the PSNR values for various methods, namely CYCLE-GAN, PALETTE, and NCDM, providing a comparative overview of their performance.

Model	PSNR (dB)	SSIM
CYCLE-GAN	25.50	0.999
Palette	24.54	0.48
NCDM (Our Approach)	37.28	0.935

TABLE 5.1: Comparative analysis of image enhancement models: CYCLE-GAN, Palette, and NCDM.

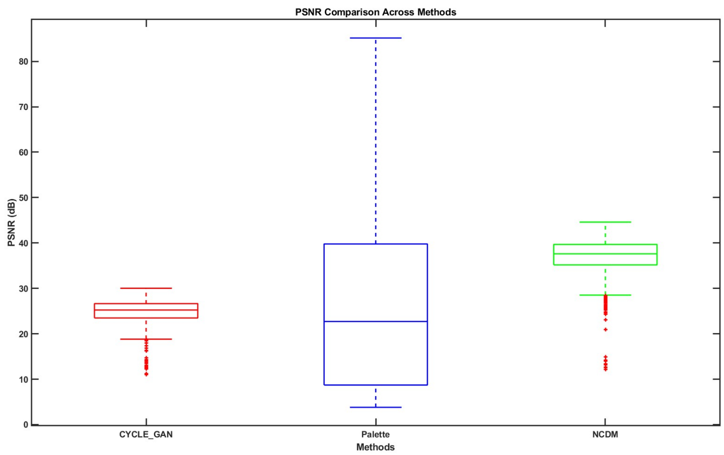

FIGURE 5.2: Comparison of PSNR values across different methods: CYCLE-GAN, Palette, and NCDM.

Figure 5.2 illustrates a boxplot analysis of PSNR values for different image enhancement methods. The CYCLE-GAN method's results with a lower median PSNR value and a substantial number of low outliers, suggesting this method can sometimes result in lower-quality enhancements. The Palette Methods, represented in the middle in blue, show a much higher median PSNR value and a wider range of values, indicating some inconsistency in image enhancement. NCDM, illustrated on the right in green, has a median PSNR that is comparable to the Palette Methods but with a narrower interquartile range, suggesting it not only improves image quality but also does so with greater consistency. In this comparison, while the Palette Methods and NCDM both demonstrate an ability to enhance image quality effectively, NCDM might be preferred for its consistency, evidenced by a tighter interquartile range, suggesting a reliable enhancement across different images.

5.2.2 SSIM Comparison

In figure 5.3, the comparative efficacy of image enhancement techniques is evaluated using the Structural Similarity Index Measure (SSIM), which assesses how similar an enhanced image is to a reference. The SSIM comparison across the CYCLE-GAN, Palette Methods, and NCDM demonstrates varying levels of structural preservation and image quality enhancement.

The Palette, despite showing improvement over the baseline, has a broad spread of SSIM scores, indicating some inconsistencies in performance. On the other hand, both CYCLE-GAN Methods and NCDM present significantly higher SSIM values, suggesting a superior capability to retain the original image structures after enhancement. Among them, CYCLE-GAN stands out for its tightly grouped SSIM scores around a high median value, suggesting that it consistently maintains structural details well.

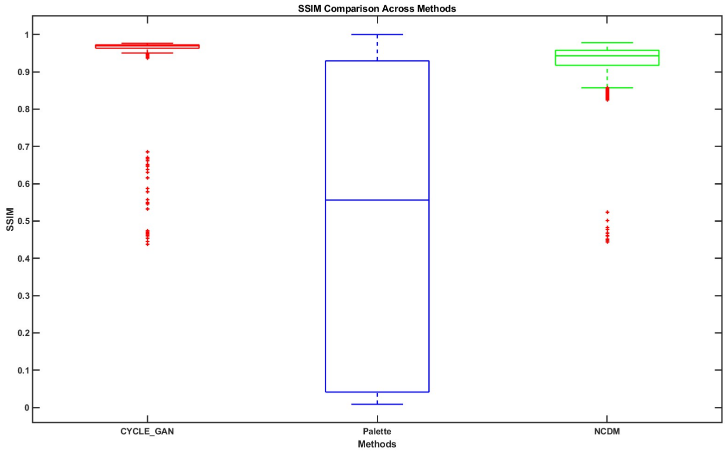

FIGURE 5.3: Comparison of SSIM values across different methods: Original, CYCLE-GAN, Palette, and NCDM.

The dual analysis underscores NCDM's robustness as a technique that consistently enhances image quality while preserving structural details. In fields where image fidelity is paramount, such as medical imaging, NCDM's predictable performance could be especially valuable. This reliability, as evidenced by its concentrated SSIM scores and narrow PSNR range, positions NCDM as a potentially superior choice for high-quality image enhancement.

5.3 Visual Quality Assessment

This section is dedicated to the visual assessment of low-dose CT (LDCT) test images enhanced by various techniques, including the CYCLE-GAN technique, the Palette method, and our proposed Noise-Characterized Diffusion Model (NCDM).

This comparison aims to evaluate the effectiveness of each method in reducing noise and preserving important image details. The visual outcomes of applying different image enhancement techniques to sample low-dose CT images are depicted as follows:

- Figure 5.4 showcases the results of the CYCLE-GAN technique, demonstrating its capability in noise reduction while preserving the anatomical structures.
- from Figure 5.5 illustrates the improvements achieved with the Palette method, showing notable advancements in image quality.
- from Figure 5.6 highlights our NCDM technique's performance, revealing a significant enhancement in image clarity with finer details and reduced noise levels.

Through these visual comparisons, we can assess the effectiveness of each method in mimicking the quality expected from high-dose imaging while using lower radiation-dose scans.

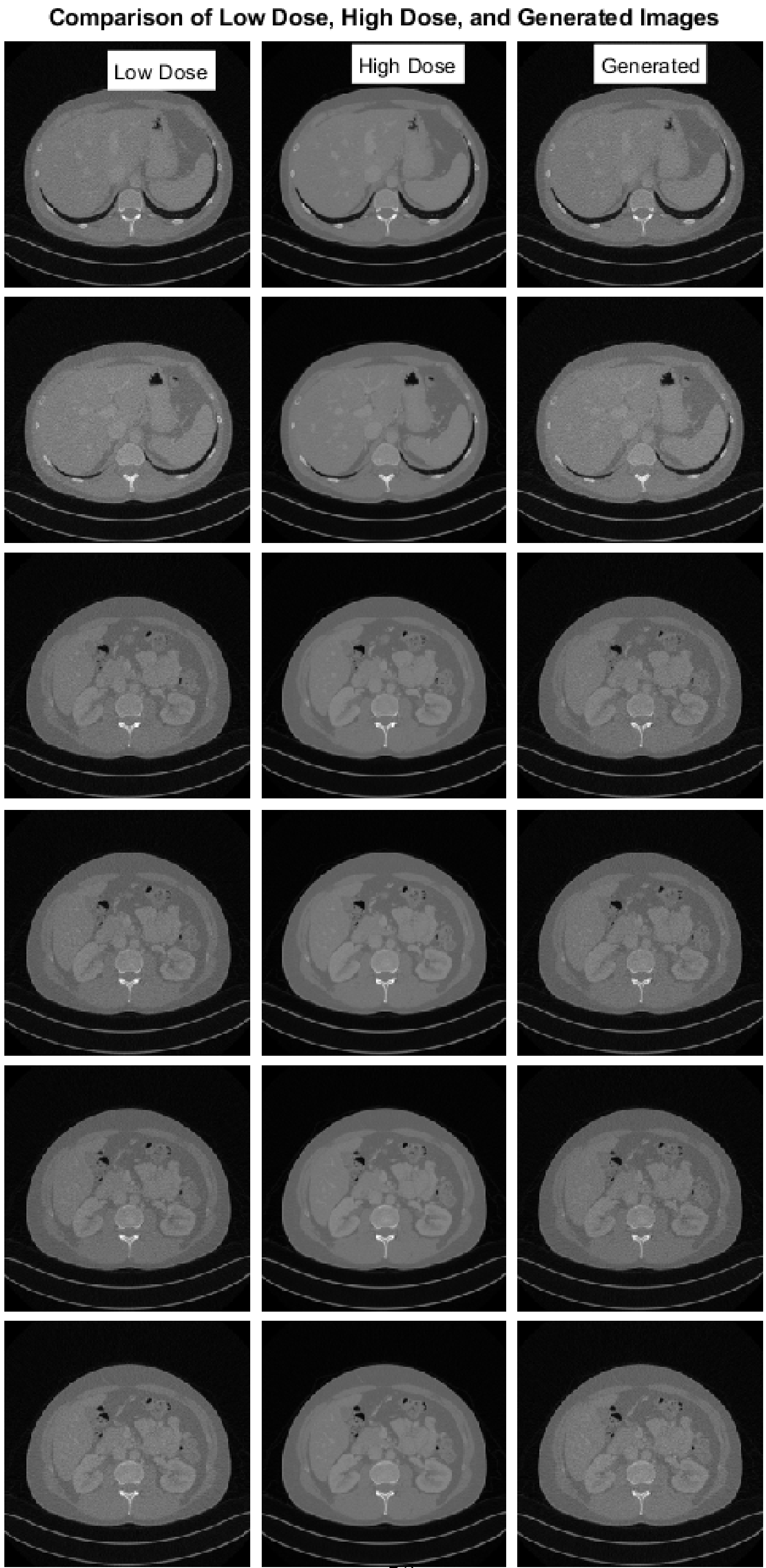

Figure 5.4: HDCT image enhanced by the CYCLE-GAN technique.

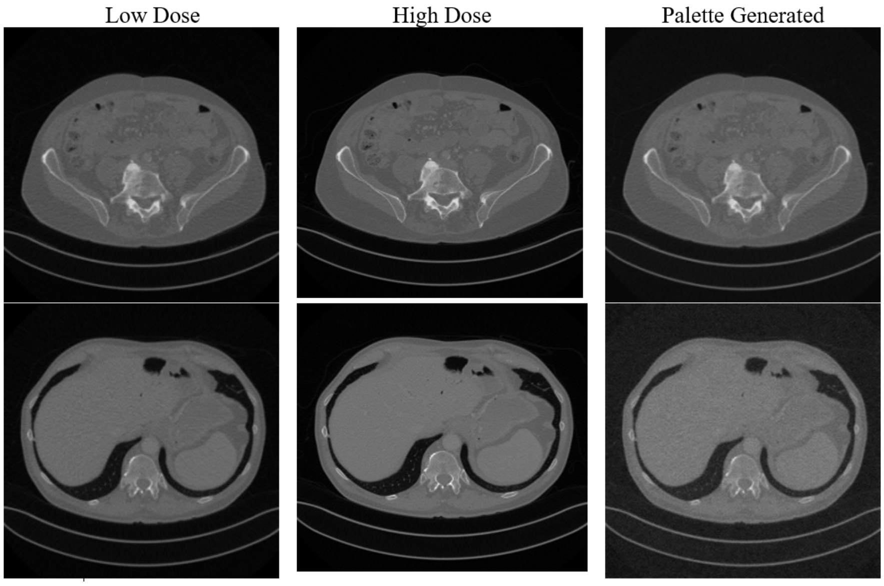

Figure 5.5: HDCT image enhanced by the Palette method.

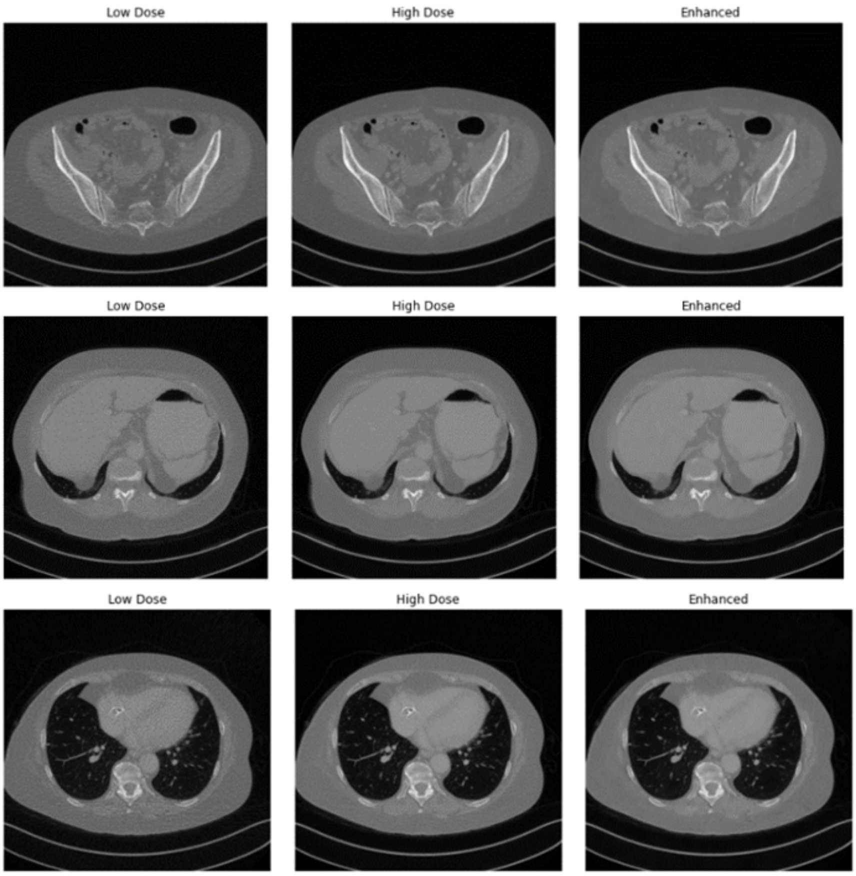

Figure 5.6: HDCT image enhanced by our NCDM.

5.4 Conclusion

The visual and quantitative analysis provided in this chapter emphasizes the effectiveness and advantages of the Noise-Characterized Diffusion Model (NCDM) over other methods, such as the CYCLE-GAN and Palette. By incorporating detailed noise characterization and advanced diffusion techniques, NCDM significantly improves the quality and diagnostic value of low-dose CT images.

Chapter 6

Conclusion and Future Work

6.1 Conclusion

This thesis explored the innovative use of generative diffusion models for enhancing the quality of low-dose CT scans. By integrating detailed noise characterization with generative diffusion processes, we developed models that improve image clarity and diagnostic utility while minimizing radiation exposure, addressing a critical concern in medical imaging.

In Chapter 4, we introduced an advanced noise-characterized diffusion model framework that adapted and extended traditional diffusion model equations to tackle unique noise patterns observed in low-dose CT images. Our approach incorporated both forward and reverse processes, meticulously aligning the model with the specific challenges presented by CT imaging noise.

The implementation of these models underscores the practical applicability and effectiveness of generative diffusion techniques in medical imaging. The strategies devised in this research exemplify how advanced computational models can be harnessed to improve patient outcomes.

6.2 Future Work

The achievements of this thesis pave the way for multiple directions for future research, notably in the optimization and application of generative diffusion models within and beyond medical imaging:

6.2.1 Model Optimization and Efficiency

Future research could focus on enhancing the computational efficiency and optimization algorithms of noise-characterized diffusion models, enabling faster processing times without compromising image quality. This includes developing more sophisticated models that can handle diverse and complex noise patterns more effectively.

6.2.2 Expansion to Other Imaging Modalities

While this thesis focused on low-dose CT scans, the methodologies developed can be adapted and applied to other imaging modalities, such as MRI or PET scans. Exploring these applications could widen the impact of diffusion models in medical imaging.

6.2.3 Machine Learning and AI Integration

There is a substantial opportunity to integrate machine learning and AI more deeply into the development and refinement of diffusion models. This could involve using deep learning to improve model training efficiency or employing reinforcement learning for automatic parameter tuning.

6.2.4 Patient-Specific Modeling

Future work could also explore the development of patient-specific diffusion models that take into account individual variations in anatomy and tissue composition, potentially leading to personalized imaging techniques that optimize image quality and diagnostic accuracy for each patient.

By addressing these areas, future research can continue to advance the field of medical imaging, leveraging the full potential of generative diffusion models to enhance diagnostic processes while maintaining patient safety.

Bibliography

Ambrose, J. (1976). Computerized transverse axial scanning (tomography): Part 2. Clinical application. *British Journal of Radiology* 49(577), 833–840.

Beister, M., Kolditz, D., and Kalender, W. A. (2012). Iterative reconstruction methods in X-ray CT. *Physica Medica* 28(2), 94–108.

Body CT (CAT Scan) (2024). `https://www.radiologyinfo.org/en/info/bodyct`. Accessed: insert date here.

Brenner, D. J. and Hall, E. J. (2007). Computed Tomography — An Increasing Source of Radiation Exposure. *The New England Journal of Medicine* 357(22), 2277–2284.

Cormack, A. M. (1973). Representation of a Function by Its Line Integrals, with Some Radiological Applications. *Journal of Applied Physics* 34(9), 2722–2727.

Costa, P., Galdran, A., Meyer, M. I., Niemeijer, M., Abramoff, M., Mendonça, A. M., and Campilho, A. (2018). End-to-end adversarial retinal image synthesis. *IEEE Transactions on Medical Imaging* 37(3), 781–791.

Dhariwal, P. and Nichol, A. (2021). Diffusion Models Beat GANs on Image Synthesis. In: *Advances in Neural Information Processing Systems*.

Dobbins, J. T., Frush, D. P., Kigongo, C. J., MacFall, J. R., Reiman, R. E., and Trahey, G. E. (2014). Medical imaging safety in the developing world. In: *Radiology in Global Health: Strategies, Implementation, and Applications*, 41–60.

Edelmers, E., Kazoka, D., Bolocko, K., Sudars, K., and Pilmane, M. (2024). Automatization of CT Annotation: Combining AI Efficiency with Expert Precision. *Diagnostics* 14(2), 185.

Flohr, T. G., McCollough, C. H., Bruder, H., Petersilka, M., Gruber, K., Süss, C., Grasruck, M., Stierstorfer, K., Krauss, B., Raupach, R., Primak, A. N., Küttner, A., Achenbach, S., Becker, C., Kopp, A., and Ohnesorge, B. M. (2005). Multislice CT: technical principles and future trends. *European Radiology* 15(D3), 23–30.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. In: *Advances in Neural Information Processing Systems*. Vol. 27, 2672–2680.

Greenspan, H., Ginneken, B. van, and Summers, R. M. (2016). Guest Editorial Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique. *IEEE Transactions on Medical Imaging* 35(5), 1153–1159.

Han, A., Other, A., and SoOn, A. (2020). Title of the Han et al. article. *Journal Name* Volume Number(Issue Number), Page Range.

Hara, A. K., Paden, R. G., Silva, A. C., Kujak, J. L., Lawder, H. J., and Pavlicek, W. (2009). Iterative reconstruction technique for reducing body radiation dose at CT: feasibility study. *American Journal of Roentgenology* 193(3), 764–771.

Ho, J., Jain, A., and Abbeel, P. (2020). Denoising Diffusion Probabilistic Models. In: *Advances in Neural Information Processing Systems*.

Hounsfield, G. N. (1973). Computerized transverse axial scanning (tomography): Part 1. Description of system. *British Journal of Radiology* 46(552), 1016–1022.

Hsieh, J. (2003). Adaptive Statistical Iterative Reconstruction: From the Basics to the Clinic. *Physics in Medicine and Biology* 48(14), R65–R88.

Hung, A. L. Y., Zhao, K., Zheng, H., Yan, R., Raman, S. S., Terzopoulos, D., and Sung, K. (2023). Med-cDiff: Conditional medical image generation with diffusion models. *Bioengineering* 10(11), 1258.

Kalender, W. A., Seissler, W., Klotz, E., and Vock, P. (1989). Spiral volumetric CT with single-breath-hold technique, continuous transport, and continuous scanner rotation. *Radiology* 176(1), 181–183.

Karimi, D., Deman, P., Ward, R., and Ford, N. (2016). A sinogram denoising algorithm for low-dose computed tomography. *BMC medical imaging* 16, 1–14.

Kawamura, M., Kamomae, T., Yanagawa, M., Kamagata, K., Fujita, S., Ueda, D., and Naganawa, S. (2024). Revolutionizing radiation therapy: the role of AI in clinical practice. *Journal of Radiation Research* 65(1), 1–9.

Kazeminia, S., Baur, C., Kuijper, A., Ginneken, B. van, Navab, N., Albarqouni, S., and Mukhopadhyay, A. (2020). GANs for medical image analysis. *Artificial Intelligence in Medicine* 109, 101938.

Kazerouni, A., Aghdam, E. K., Heidari, M., Azad, R., Fayyaz, M., Hacihaliloglu, I., and Merhof, D. (2023). Diffusion Models in Medical Imaging: A Comprehensive Survey. *Medical Image Analysis*, 102846.

Kim, B. and Seo, Y. (2023). A Study of Pattern Defect Data Augmentation with Image Generation Model. *Journal of the Korea Computer Graphics Society* 29(3), 79–84.

Li, K., Tang, J., and Chen, G.-H. (2014). Iterative Reconstruction for CT Imaging. *Journal of X-Ray Science and Technology* 22(3), 393–420.

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Laak, J. A. van der, Ginneken, B. van, and Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. In: *Medical image analysis*. Vol. 42. Elsevier, 60–88.

MathWorks (2023). *Unsupervised Medical Image Denoising Using CycleGAN*. `https://www.mathworks.com/help/images/unsupervised-medical-image-denoising-using-cyclegan.html`. Accessed: 2024-04-08.

McCollough, C., Chen, B., Holmes III, D. R., Duan, X., Yu, Z., Yu, L., Leng, S., and Fletcher, J. (2020). Low Dose CT Image and Projection Data (LDCT-and-Projection-Data). *The Cancer Imaging Archive* 2020(1), 1–10.

Miller, D. L. and Schauer, D. (1983). The ALARA principle in medical imaging. *Philosophy* 44(1), 595–600.

Mirsky, Y. and Lee, W. (2021). The creation and detection of deepfakes: A survey. *ACM Computing Surveys (CSUR)* 54(1), 1–41.

Pearce, M. S., Salotti, J. A., Little, M. P., McHugh, K., Lee, C., Kim, K. P., Howe, N. L., Ronckers, C. M., Rajaraman, P., Craft, A. W., Parker, L., and González, A. B. de (2012). Radiation exposure from CT scans in childhood and subsequent risk of leukaemia and brain tumours: a retrospective cohort study. *The Lancet* 380(9840), 499–505.

Pianykh, O. S. (2020). Digital Imaging and Communications in Medicine (DICOM): A Practical Introduction and Survival Guide. *Journal of Digital Imaging* 33(1), 279–289.

Pugliesi, R. A. (2018). Recent Developments in AI Algorithms for Pediatric Radiology: Advancements in Detection, Diagnosis, and Management. *International Journal of Applied Health Care Analytics* 3(10), 1–20.

Ramirez Giraldo, J. C., Trzasko, J. D., Leng, S., and McCollough, C. H. (2011). Iterative Reconstruction Techniques in CT: Technical Principles and Clinical Applications. *RadioGraphics* 31(5), 1483–1503.

Rueckert, D. and Schnabel, J. A. (2019). Model-based and data-driven strategies in medical image computing. *Proceedings of the IEEE* 108(1), 110–124.

Saharia, C. et al. (2022). Palette: Image-to-image diffusion models. In: *ACM SIGGRAPH 2022 conference proceedings*.

Schauer, D. A. and Linton, O. W. (2009). The ALARA concept in pediatric CT: myth or reality? *Radiology* 252(3), 603–607.

Shen, D., Wu, G., and Suk, H.-I. (2017). Deep Learning in Medical Image Analysis. *Annual Review of Biomedical Engineering* 19, 221–248.

Smith, N. B. and Webb, A. (2010). *Introduction to medical imaging: physics, engineering and clinical applications*. Cambridge University Press.

Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., and Ganguli, S. (2015). Deep Unsupervised Learning using Nonequilibrium Thermodynamics. *arXiv preprint arXiv:1503.03585*.

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B. (2020). Score-Based Generative Modeling through Stochastic Differential Equations. *arXiv preprint arXiv:2011.13456*.

Wang, D., Khosla, A., Gargeya, R., Irshad, H., and Beck, A. H. (2017). Deep Learning for Identifying Metastatic Breast Cancer. *arXiv preprint arXiv:1606.05718*.

Wang, S., Yang, D. M., Rong, R., Zhan, X., and Xiao, G. (2019). Pathology image analysis using segmentation deep learning algorithms. *The American Journal of Pathology* 189(9), 1686–1698.

Willemink, M. J. and Noël, P. B. (2019). The evolution of image reconstruction for CT—from filtered back projection to artificial intelligence. *European Radiology* 29, 2185–2195.

Yamashita, R., Nishio, M., Do, R. K. G., and Togashi, K. (2018). Convolutional Neural Networks: An Overview and Application in Radiology. *Insights into Imaging* 9(4), 611–629.

Yang, L., Zhang, Z., Song, Y., Hong, S., Xu, R., Zhao, Y., and Yang, M. H. (2023). Diffusion Models: A Comprehensive Survey of Methods and Applications. *ACM Computing Surveys* 56(4), 1–39.

Yi, X., Walia, E., and Babyn, P. (2019). Generative adversarial network in medical imaging: A review. *Medical Image Analysis* 58, 101552.

Yim, J., Stärk, H., Corso, G., Jing, B., Barzilay, R., and Jaakkola, T. S. (2024). Diffusion models in protein structure and docking. *Wiley Interdisciplinary Reviews: Computational Molecular Science* 14(2), e1711.

You, C. et al. (2018). Structurally-sensitive multi-scale deep neural network for low-dose CT denoising. *IEEE Access* 6, 41839–41855.