UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Francisco Mauro Falcão Matias Filho

A Power-Law Noise Scheduler for Image Diffusion Models in Low-Dose CT Denoising

Recife

2025

Francisco Mauro Falcão Matias Filho

A Power-Law Noise Scheduler for Image Diffusion Models in Low-Dose CT Denoising

Dissertação apresentada ao Programa de Pós-graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco, como requisito parcial para obtenção do grau de Mestre em Ciência da Computação.

Área de Concentração: Inteligência Computacional

Orientador: Prof. Dr. George Darmiton da Cunha Cavalcanti

Coorientador: Prof. Dr. Luis Filipe Alves Pereira

Recife

2025

.Catalogação de Publicação na Fonte. UFPE - Biblioteca Central

Matias Filho, Francisco Mauro Falcão.
A Power-Law noise scheduler for image diffusion models in low-dose CT denoising / Francisco Mauro Falcão Matias Filho. - Recife, 2025.
94f.: il.

Dissertação(Mestrado)- Universidade Federal de Pernambuco, Centro de Informática, Programa de Pós-Graduação em Ciência da Computação, 2025.
Orientação: George Darmiton da Cunha Cavalcanti.
Coorientação: Luis Filipe Alves Pereira.

1. Tomografia computadorizada de baixa dose; 2. Remoção de ruído; 3. Redes de difusão; 4. Rotina de adição de ruído; 5. Aprendizado por currículo. I. Cavalcanti, George Darmiton da Cunha. II. Pereira, Luis Filipe Alves. III. Título.

UFPE-Biblioteca Central

Francisco Mauro Falcão Matias Filho

“A Power-Law Noise Scheduler for Image Diffusion Models in Low-Dose CT Denoising”

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Ciência da Computação. Área de Concentração: Inteligência Computacional;

Aprovado em: 31/07/2025.

BANCA EXAMINADORA

_____

Prof. Dr. Cleber Zanchettin
Centro de Informática / UFPE

_____

Prof. Dr. David Menotti Gomes
Departamento de Informática / UFPR

_____

Prof. Dr. George Darmiton da Cunha Cavalcanti
Centro de Informática/UFPE
(**orientador**)

To my past self, for not giving up

ACKNOWLEDGEMENTS

I begin by expressing my deepest gratitude and love to my family. To my parents, Marta and Mauro, for their constant support and encouragement throughout this journey. And to my sisters, Manuelle and Mônica, whose strength and determination have always inspired me.

I am thankful to my advisors for their guidance and mentorship during this process. To Professor George and Professor Tsang, for showing me how to approach research through their own example. And especially to Professor Luis, for the regular meetings, the constructive feedback, and for his patience in trying to understand both my reasoning and my occasional chaotic ideas.

To my beloved friends, Raquel, Cecé, Lais, Abner, Evan, Luan, and Bete, thank you for the moments we shared, the distractions, and your unconditional friendship. This journey would not have been possible without you. You are, and always will be, my chosen family. To the friends of friends who let me tag along when I was too shy to show up on my own. Thank you for your kindness and openness.

To my fellow master's students Lívia and Marcela, thank you for the shared study sessions and the memes that helped keep the balance. To my colleagues from the LuxAI project, thank you for the collaboration and support. You have been a quiet but constant inspiration to become a better researcher.

To the professors at UFC, who played an essential role at the beginning of my academic path. I am especially grateful to Professor Iális and Professor Jermana, who supported me early on, when I was still learning to believe in myself. I also thank UFPE and CIn for the infrastructure and academic environment that made this work possible. And to CAPES, for the financial scholarship that supported my studies and research.

Finally, to all those who believe in the value of public, inclusive, and high-quality higher education, accessible to everyone, regardless of gender, race, background, or orientation.

"Quando você escolhe o novo,

o velho tenta ficar.

Não aceite.

Já partiu."

(Autor desconhecido)

ABSTRACT

The growing need to reduce patient exposure to ionizing radiation in medical imaging has led to the widespread adoption of low-dose computed tomography protocols, guided by the principle of ALARA (As Low As Reasonably Achievable). However, lowering the radiation dose results in increased image noise and artifacts, which can significantly compromise diagnostic quality. Advanced deep learning techniques have demonstrated that neural networks can effectively learn mappings between LDCT and corresponding NDCT images, improving reconstruction quality. Nevertheless, most existing approaches rely on direct mappings, which may have reached a performance plateau, as suggested by the diminishing improvements observed across recent models. In this context, diffusion models have emerged as promising alternatives for image restoration due to their generative capabilities. However, traditional formulations based on stochastic noise addition may not be well-suited when modeling complex noise distributions. Previous works have addressed this issue by redefining the forward process as a deterministic transformation between clean and noisy images, aligning more closely with actual real-world scenarios. Building upon this alternative formulation, we introduce a flexible power-law noise scheduler parameterized by an exponent γ, which controls the rate at which noise is introduced during the diffusion process. This design enables the exploration of different noise progression dynamics, and allows for the evaluation of Curriculum Learning hypotheses in the context of LDCT denoising. When $\gamma > 1$, the model follows a curriculum that gradually increases noise complexity, while $\gamma < 1$ corresponds to a reverse curriculum. This flexibility positions γ as a key component for guiding the learning process. Experimental results show that the proposed method outperforms standard reconstruction techniques, especially under aggressive noise schedules. Additionally, the best results were obtained with $\gamma > 1$, reinforcing the effectiveness of a curriculum-based degradation strategy in LDCT reconstruction tasks.

Keywords: Low Dose CT. Medical Image Denoising. Diffusion Models. Noise Schedule. Curriculum Learning.

RESUMO

A crescente necessidade de reduzir a exposição dos pacientes à radiação ionizante em imagens médicas levou à adoção generalizada de protocolos de tomografia computadorizada de baixa dose, guiados pelo princípio ALARA (Tão Baixo Quanto Razoavelmente Exequível). No entanto, a redução da dose de radiação resulta em aumento do ruído e artefatos nas imagens, o que pode comprometer significativamente a qualidade diagnóstica. Técnicas avançadas de deep learning têm demonstrado que redes neurais podem aprender efetivamente mapeamentos entre imagens LDCT e as correspondentes NDCT, melhorando a qualidade da reconstrução. Contudo, a maioria das abordagens existentes baseia-se em mapeamentos diretos, que podem ter atingido um platô de desempenho, como sugerido pelas melhorias cada vez menores observadas em modelos recentes. Nesse contexto, modelos de difusão surgem como alternativas promissoras para restauração de imagens devido às suas capacidades generativas. Entretanto, formulações tradicionais baseadas na adição estocástica de ruído podem não ser adequadas para modelar distribuições complexas de ruído. Trabalhos anteriores abordaram essa questão redefinindo o processo de adição de ruído como uma transformação determinística entre imagens limpas e ruidosas, alinhando-se mais de perto com cenários reais. Com base nessa formulação alternativa, esta dissertação propõe uma rotina de adição de ruído flexível baseada em uma lei de potência, parametrizada por um expoente γ, que controla a taxa em qual o ruído é introduzido durante o processo de difusão. Esse design possibilita a exploração de diferentes dinâmicas de progressão do ruído e permite a avaliação de hipóteses de Aprendizado por Currículo no contexto de remoção de ruído em imagens médicas. Quando $\gamma > 1$, o modelo segue um currículo que aumenta gradualmente a complexidade do ruído, enquanto $\gamma < 1$ corresponde a um currículo reverso. Essa flexibilidade posiciona γ como um componente chave para guiar o processo de aprendizado. Resultados experimentais mostram que o método proposto supera técnicas padrão de reconstrução, especialmente sob agendamentos de ruído mais agressivos. Além disso, os melhores resultados foram obtidos com $\gamma > 1$, reforçando a eficácia de uma estratégia de degradação baseada em currículo nas tarefas de reconstrução de tomografias de baixa dose.

Palavras-chaves: TC de Baixa Dose. Remoção de ruído. Redes de Difusão. Rotina de Adição de Ruído. Aprendizado por currículo.

LIST OF FIGURES

LIST OF ALGORITHMS

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

AAPM	American Association of Physicists in Medicine
CNN	Convolutional Neural Network
CT	Computed Tomography
DDPM	Denoising Diffusion Probabilistic Model
DPM	Diffusion Probabilistic Model
GAN	Generative Adversarial Network
LDCT	Low-Dose Computed Tomography
NDCT	Normal-Dose Computed Tomography
NIBIB	National Institute of Biomedical Imaging and Bioengineering
PSNR	Peak Signal-to-Noise Ratio
RED-CNN	Residual Encoder-Decoder Convolutional Neural Network
RMSE	Root Mean Squared Error
SACNN	Self-Attention Convolutional Neural Network
SNR	Signal-to-Noise Ratio
SSIM	Structure Similarity Index Measure
VAE	Variational AutoEncoder
WGAN	Wasserstein Generative Adversarial Network

LIST OF SYMBOLS

$\mathbf{x}_0$	Input image (ground truth)
$\mathbf{x}_t$	Latent state at timestep t in the diffusion process
$\mathbf{x}_T$	Final latent variable in the forward process
t	Timestep index in the diffusion process
T	Total number of diffusion steps
β_t	Noise schedule coefficient at timestep t
α_t	Signal preservation coefficient
$\bar{\alpha}_t$	Cumulative product: $\prod_{i=1}^{t} \alpha_i$
$\boldsymbol{\epsilon}$	Gaussian noise $\sim \mathcal{N}(0, \mathbf{I})$
μ	Mean of a Gaussian distribution
Σ	Covariance of a Gaussian distribution
θ	Parameters of the neural network
$\mathcal{L}$	Loss function
D_{KL}	Kullback-Leibler divergence
$\mathcal{N}$	Gaussian distribution
$\mathcal{N}(\mu, \Sigma)$	Gaussian distribution with mean μ and covariance Σ
$\mathbf{y}$	Conditioning
γ	Power-law modulation factor for the noise schedule
$\boldsymbol{\epsilon}_\theta$	Noise prediction network
$\mathbf{U}_\theta$	Image prediction network
∇_θ	Gradient with respect to θ
$\mathbb{E}$	Expected value

D_{KL}	Kullback–Leibler divergence
$\propto$	Proportional to
$q(\mathbf{x}_0)$	Data distribution
$q(\mathbf{x}_t)$	Unconditioned marginal distribution of the noisy image at timestep t
$q(\mathbf{x}_T)$	Prior distribution at final step T, ideally close to pure Gaussian noise
$q(\mathbf{x}_t \mid \mathbf{x}_{t-1})$	Conditional probability of the forward process
$q(\mathbf{x}_{1:T} \mid \mathbf{x}_0)$	Joint distribution of the forward process
$q(\mathbf{x}_t \mid \mathbf{x}_0)$	Marginal distribution of $\mathbf{x}_t$ conditioned directly on the original image $\mathbf{x}_0$
$q(\mathbf{x}_0, \mathbf{x}_t)$	Joint distribution of the clean and noisy image at timestep t under the forward process
$q(\mathbf{x}_{t-1} \mid \mathbf{x}_t)$	Theoretical reverse of the forward process
$p_\theta(\mathbf{x}_{0:T})$	Joint distribution of the learned reverse process, from noise to clean image
$p_\theta(\mathbf{x}_{t-1} \mid \mathbf{x}_t)$	Learned reverse transition distribution from $\mathbf{x}_t$ to $\mathbf{x}_{t-1}$

CONTENTS

1 INTRODUCTION

This initial chapter introduces the research context of this master's thesis, providing an overview of denoising algorithms for medical imaging, and discusses the motivations that justify its development. The chapter also defines the main goal and specific objectives of the study, highlighting the key contributions achieved. Finally, it provides an overview of the document structure.

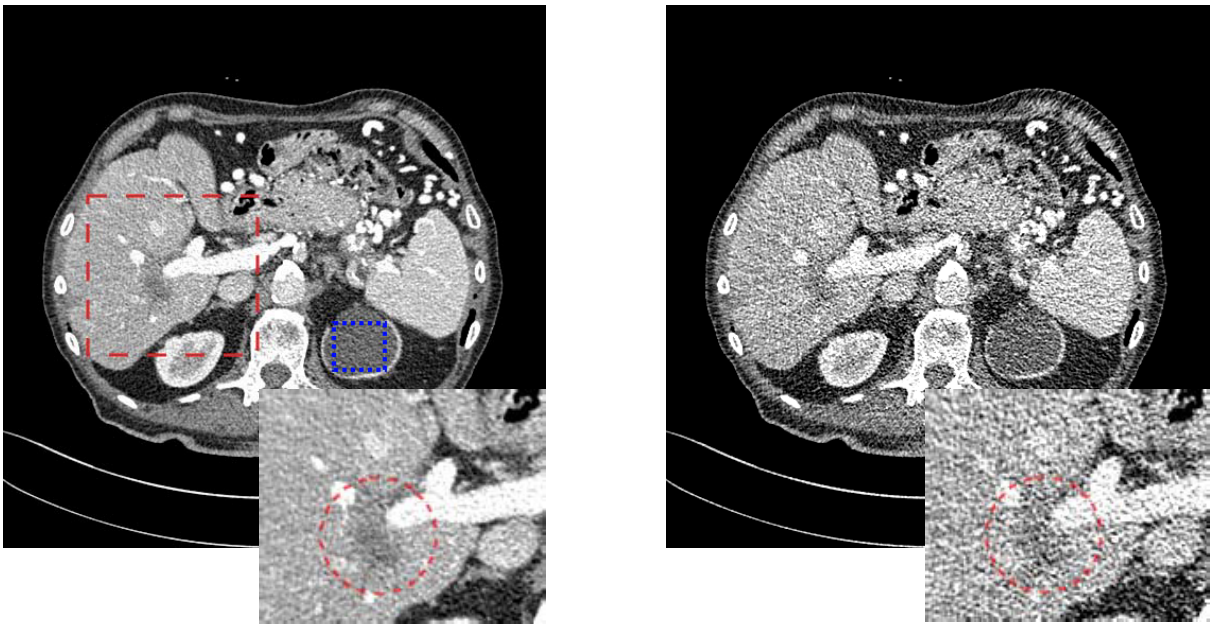
1.1 CONTEXT

Computed Tomography (CT) is a diagnostic medical imaging procedure that uses a combination of X-rays to produce detailed visualizations of internal body parts such as bones, muscles, organs, and blood vessels. Unlike standard X-ray, which uses a single beam of energy aimed directly at the desired body part, in a CT scan a series of rays are rotated around a specific body part, and a computed generated cross-sectional image is produced (PATEL; JESUS, 2023). This method is widely used in clinical procedures because, in addition to its non-invasive nature, the images produced can help to detect several pathological abnormalities, including tumors, nodules, vascular problems, and fractures (METTLER et al., 2000).

Although CT can provide an excellent clinicopathological correlation for a suspected illness (KISTLER et al., 1975), its use involves ionizing radiation, which has the potential to be carcinogen depending on factors such as the type of radiation, exposure time, irradiated tissue and age (KRILLE et al., 2010). Of all X-ray-based diagnostics, CT scans constitute one of the highest radiation dose-demanding (KALRA et al., 2002), requiring from 50 to 1000 times the radiation dose than conventional X-rays (PATEL; JESUS, 2023). This overexposure can lead to the development of metabolic abnormalities, genetic disorders, and even cancer, which can significantly reduce the patient's quality of life (MEULEPAS et al., 2019).

To address this issue, Low-Dose Computed Tomography (LDCT) protocols have emerged as an alternative to minimize radiation exposure while trying not to affect clinical dignosis. Usually, to obtain LDCT images, the X-ray flux is deliberately reduced during the procedure by lowering the tube voltage, which lowers the tube current, and effectively reduces the amount of radiation that penetrates the body (RAMPINELLI; ORIGGI; BELLOMI, 2013). However, reducing the radiation dose of the X-rays affects the Signal-to-Noise Ratio (SNR) of the X-ray signals

and results in low contrast CT images with noise and artifacts, as well as blurred edges and low textures (KULATHILAKE et al., 2023), which reduces the confiability of the diagnosis. Figure 1 shows the quality contrast between a normal-dose and a low-dose CT.



(a) NDCT (b) LDCT

Figure 1 – Comparison between normal-dose (a) and low-dose (b) abdomen CT images show a liver lesion (red box) and a cystic lesion in the left kidney (blue box). The red circle represents metastasis, which is difficult to identify in the LDCT image. (**Source:** Yang et al. (2018)).

To address this challenge, several denoising algorithms have been proposed to recover visual details in LDCT images, in order to improve their clinical usability. These methods aim to reduce the quality difference between LDCT and Normal-Dose Computed Tomography (NDCT) images. Over the years, machine learning and deep learning techniques have become widely adopted in CT protocols (ZHANG; GU; SHI, 2022).

The classical deep learning-based approach to LDCT denoising consists of learning a function f in the image space that maps a noisy LDCT image to its corresponding clean NDCT counterpart. This supervised mapping framework has led to substantial performance improvements and has been explored in a wide range of architectures, including Convolutional Neural Network (CNN), encoder-decoder structures, and residual learning models.

However, recent comparative studies such as that by Eulig et al. (2024) suggest that the performance gains between the most recent models are becoming marginal when evaluated using standard quantitative metrics. This observation raises important questions about the potential saturation of performance in direct mapping approaches, especially in scenarios where the noise characteristics are complex and non-Gaussian, as is the case with real-world LDCT

data. In this context, generative models have emerged as powerful frameworks for learning data distributions and generating high-quality synthetic samples.

Generative models are a class of deep learning models designed to learn the underlying data distribution of a given dataset, enabling the generation of new samples that are statistically similar to the training data. Formally, these models are trained to learn the probability distribution $p_{data}(\mathbf{x})$ of given dataset and sample new data $\hat{\mathbf{x}} \sim p_\theta(\mathbf{x})$, where θ represents the learnable parameters of the learned model (RUTHOTTO; HABER, 2021). In the context of imaging, samples are typically represented as two-dimensional matrices $\mathbf{x} \in \mathbb{R}^{H \times W}$ or, alternatively, as high-dimensional vectors obtained by flattening $H \times W$ images into one-dimensional sequences (BISHOP, 1995).

Traditional generative models include Generative Adversarial Network (GAN) (GOODFELLOW et al., 2014), Variational AutoEncoder (VAE) (KINGMA; WELLING, 2013), and Flow-based models (DINH; KRUEGER; BENGIO, 2014; DINH; SOHL-DICKSTEIN; BENGIO, 2016; KINGMA; DHARIWAL, 2018). They have shown great success in high-quality sampling generation, but each method has its own limitations. GANs are known for training instability due to their adversarial nature, and can easily suffer from mode collapse (THANH-TUNG; TRAN, 2020), which affects sampling diversity. VAEs are more stable, but often produce blurry images and lack detail (HUANG et al., 2018). Flow models require specialized reversible architectures that are computationally expensive (CHEN et al., 2019).

More recently, the Denoising Diffusion Probabilistic Model (DDPM) (HO; JAIN; ABBEEL, 2020) has emerged as a new class of latent variable generative models that works by gradually adding noise to an input signal, and then learns to reverse this process in order to generate new samples. Figure 2 shows a comparative scheme between these four generative models (WENG, 2021).

In the field of generative imaging, denoising diffusion models became the state of the art in many tasks, including text-to-image generation, inpainting, super-resolution, and image restoration (NICHOL et al., 2021; NICHOL; DHARIWAL, 2021; RAMESH et al., 2021; MENG et al., 2022), having demonstrated superiority over traditional approaches (DHARIWAL; NICHOL, 2021). However, despite their widespread use in other contexts, the application of diffusion models to image denoising remains relatively unexplored (HUNG et al., 2023; LIU et al., 2023).

This gap can be attributed to the inherit way these models operate: they use two Markov chains parameterized by a neural network. In the forward chain, noise is progressively sampled from a predefined isotropic distribution and added to the image, ultimately yielding a fully

noisy image at the end of the process. In the reverse chain, a neural network is trained to iteratively remove this noise and reconstruct the original image (HO; JAIN; ABBEEL, 2020).

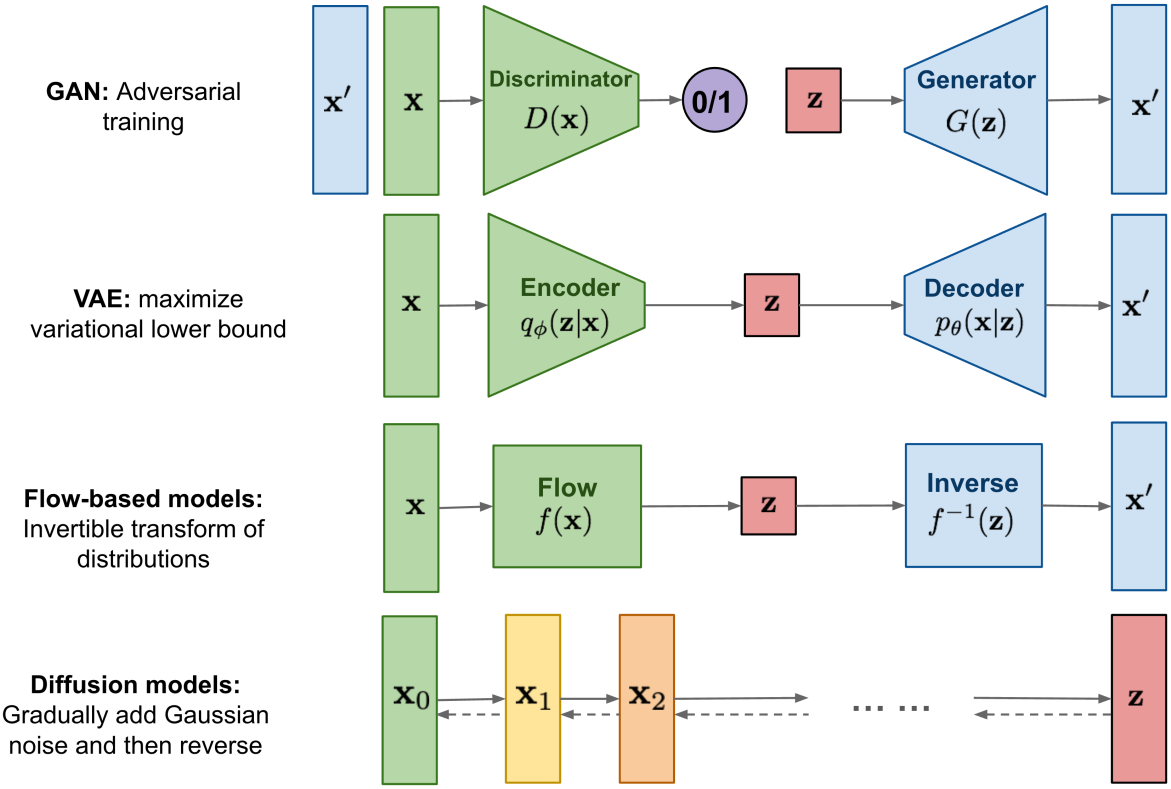


Figure 2 – Comparison between the schematic diagrams of GAN, VAE, flow model and diffusion model. Unlike the aforementioned models, DDPM is learned through an iterative procedure and the latent variable has high dimensionality, same as the input data. (**Source:** Weng (2021)).

Diffusion-based methods often perform image denoising as a conditional image generation task. In classic conditional diffusion models, the reverse process begins with pure noise and is guided by an input signal (SAHARIA et al., 2022a), which can be a class label, a text representation, a semantic map, or even, in the case of LDCT denoising, an image (XIA; LYU; WANG, 2022). This conditioning signal may be concatenated within a Convolutional Neural Network architecture along with the data input the timestep embeddings (ROMBACH et al., 2022). Figure 3 provides a general overview of the conditioning process for image denoising.

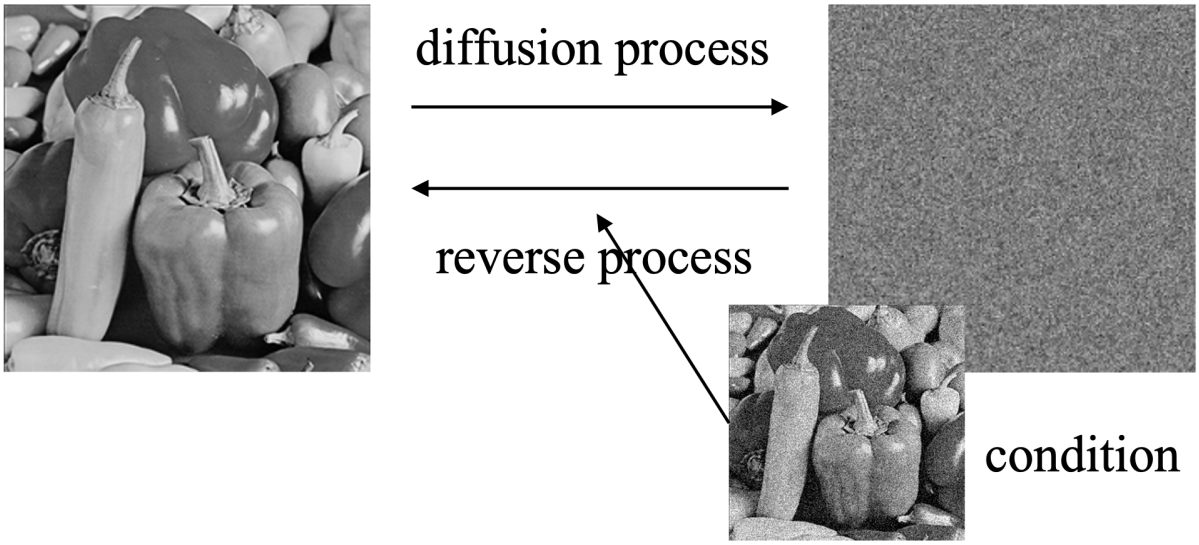


Figure 3 – Illustration of the original diffusion model approach applied to image denoising. The forward process remains the same: a certain amount of noise is gradually added until the image becomes pure noise. In the reverse process, the conditioning image is embedded into the network along with the timestep embeddings. (**Source:**Xie et al. (2023)).

1.2 MOTIVATION

A key limitation in applying DDPMs to image denoising tasks lies in the mismatch between the noise distribution assumed during training and the actual noise characteristics of real-world data. In standard DDPMs, the forward process injects Gaussian noise into clean images to simulate a degradation trajectory. However, in many practical scenarios, the noise pattern is more complex and does not follow a Gaussian distribution. This discrepancy results in suboptimal performance, as the model learns to remove a type of noise that is statistically different from what is actually present in the images.

Direct Low-Dose to Normal-Dose CT mapping methods appear to have reached a plateau, as they are advancing with relatively incremental image quality gains. This trend points to a possible exhaustion of traditional approach-based solutions. Despite the aforementioned challenges, diffusion models possess an inherent advantage for denoising: their progressive refinement mechanism. By decomposing the LDCT-to-NDCT mapping into multiple denoising steps, they avoid the limitations of single-step reconstruction.

With this in mind, our methods builds upon a previously explored alternative forward process, in which the trajectory from clean NDCT images to noisy LDCT images is explicitly defined as a deterministic transition. This eliminates the need to model the noise with a predefined probability distribution and allows better alignment with the empirical noise characteristics of the data.

Additionally, we hypothesize that explicitly modeling the noise level added at each diffusion step according to a power-law schedule enables the investigation of learning schemes inspired by Curriculum Learning, where early steps involve minor degradations, or Reverse Curriculum Learning, which starts with higher noise increments.

This formulation allows us to explore how explicit difficulty modulation, interpreted here as the amount of residual noise added between successive steps, can influence the quality of LDCT-to-NDCT reconstruction and the behavior of the model during training.

1.3 OBJECTIVES

This work proposes a novel denoising strategy based on conditional diffusion models designed for low-dose CT image reconstruction. Motivated by the limitation of classical DDPMs in modeling the noise characteristics of real-world data, our primary goal is to introduce a flexible power-law noise scheduler that enables the exploration of different strategies for difficulty modulation throughout the diffusion process. In order to achieve this goal, our specific objectives are listed as follows:

- To develop a mathematical formulation for the noise scheduler based on a power-law function, parameterizing its progression through the exponent γ;
- To implement the DDPM diffusion process using multiple steps of residual noise addition, modulating the incremental noise level between steps;
- To empirically evaluate the impact of different noise scheduler configurations on LDCT-to-NDCT image reconstruction performance;
- To investigate the feasibility of interpretations based on Curriculum Learning and Reverse Curriculum Learning through residual noise modulation;
- To compare the results obtained with traditional direct mapping approaches and other well-established noise schedules in the diffusion model literature.

1.4 CONTRIBUTIONS

The following contributions are expected to be achieved upon the completion of the aforementioned objectives:

- The proposal of a novel noise scheduler compatible with the DDPM framework, based on a power-law formulation with flexible parameterization, whose approach allows control over the progression of residual noise added at each diffusion step through the hyperparameter γ, which regulates the rate of noise addition;
- An ablation study that explores how different noise scheduler configurations can modulate the incremental difficulty between successive diffusion steps;
- An analysis on the impact of explicit difficulty modulation on the steps of the image denoising learning process, introducing for the first time in the literature the discussion of Curriculum Learning and Reverse Curriculum Learning concepts applied in the context of diffusion models.

1.5 DOCUMENT STRUCTURE

This work is structured into the following chapters:

- **Chapter 2:** Presents the theoretical foundations of diffusion models, including the forward and reverse processes, and their application to image generation tasks.
- **Chapter 3:** Introduces the conditional aspects of guided diffusion models and reviews recent advances in image-conditioned diffusion, as well as their applicability to image denoising tasks.
- **Chapter 4:** Describes the proposed forward process, the conditioning strategy, the noise scheduling mechanism, and the deterministic reverse process, detailing the architectural and algorithmic choices.
- **Chapter 5:** Presents the overall experimental pipeline, detailing dataset descriptions, training protocols, and evaluation metrics.
- **Chapter 6:** Describes experimental results, including comparisons with baseline methods and ablation studies.
- **Chapter 7:** Summarizes the main findings, discusses limitations, and outlines directions for future research.

2 DENOISING DIFFUSION PROBABILISTIC MODELS

This chapter introduces the theoretical foundations of diffusion models and presents the methods and techniques on which they are based. Section 2.1 provides a brief overview of the general diffusion formulation, while Sections 2.1.1 and 2.1.2 formalize its two main components: the forward and reverse processes. Sections 2.2 and 2.3 present the training objective and model architecture proposed in the original DDPM formulation, respectively. Finally, Section 2.4 discusses the design details of noise schedules.

2.1 DIFFUSION PROCESS

The concept of diffusion originates from the field of thermodynamics, where it describes the spontaneous movement of particles from regions of higher concentrations to regions of lower concentrations driven by the principle of entropy (JARZYNSKI, 2012). In statistics, it refers to the process of transforming complex distributions simpler ones (NEAL, 1998). This intuition inspired the development of the Diffusion Probabilistic Model (DPM) (SOHL-DICKSTEIN et al., 2015), a class of generative models that also borrow principles from non-equilibrium statistical physics (JARZYNSKI, 1997) and sequential Monte Carlo methods (NEAL, 1998).

In the context of generative modeling, DPMs are considered implicit models: they learn to generate samples without explicitly defining a probability distribution over the data (WU; GAO; ZHA, 2021). Instead, they provide a way of interacting with the probability distribution indirectly. The learning process involves a forward diffusion stage, modeled as a Markov chain, where structured data is progressively corrupted with noise. A neural network is then trained to reverse this process, step by step, in order to recover the original data distribution (SOHL-DICKSTEIN et al., 2015). Figure 4 illustrates this process (KREIS; GAO; VAHDAT, 2022).

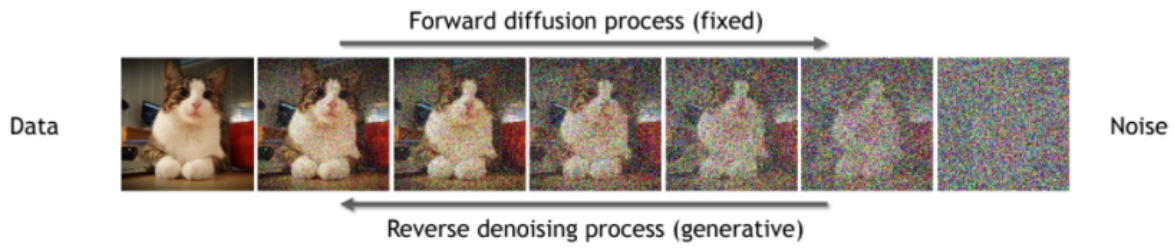


Figure 4 – General overview of the diffusion framework. The process consists of two stages: a forward diffusion process that gradually adds noise to the input, and a reverse denoising process that learns to generate data samples from noise. (**Source:** Kreis, Gao e Vahdat (2022)).

2.1.1 Forward diffusion process

In this stage, the input data $\mathbf{x}_0 \in \mathbb{R}^{H \times W}$, assumed to be continuous over pixel values to enable the diffusion modeling, sampled from the original data distribution $q(\mathbf{x}_0)$ is progressively transformed into a latent variable $\mathbf{x}_T \in \mathbb{R}^{H \times W}$ of the same dimensionality through a forward diffusion process, which consists of a Markov chain that gradually adds Gaussian noise to $\mathbf{x}_0$ over T steps, generating a sequence $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_T$, where $\mathbf{x}_T$ asymptotically approaches a standard Gaussian distribution, $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, and each intermediate step $\mathbf{x}_t \in \mathbb{R}^{H \times W}$ becomes increasingly noisier. Figure 5 illustrates this process.

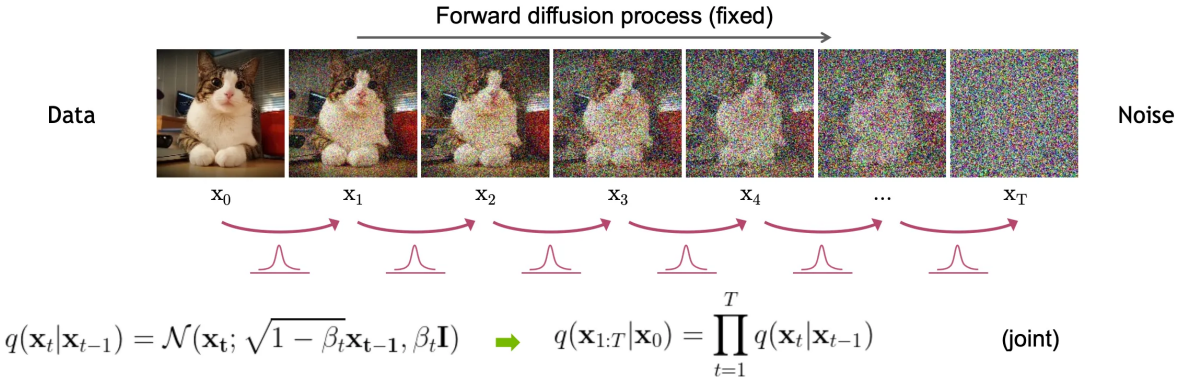


Figure 5 – Illustration of the formal definition of the forward process in T steps. The data sample $\mathbf{x}_0$ gradually loses its distinguishable features as the step t becomes larger, and eventually when $T \to \infty$, $\mathbf{x}_0$ becomes equivalent to an isotropic Gaussian distribution. (**Source:** Kreis, Gao e Vahdat (2022)).

Formally, the forward process is defined as Equation 2.1 (HO; JAIN; ABBEEL, 2020)

$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}\left(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I}\right), \tag{2.1}$$

where $q(\mathbf{x}_t|\mathbf{x}_{t-1})$ is the conditional distribution of the latent variable $\mathbf{x}_t$ given the previous state $\mathbf{x}_{t-1}$, and $\{\beta_t \in (0,1)\}_{t=1}^{T}$ is the variance schedule that controls the addition of Gaussian noise that produces the sequence $\mathbf{x}_1, \cdots, \mathbf{x}_T$. The notation $\mathcal{N}(\mathbf{x}_t; \boldsymbol{\mu}_t, \boldsymbol{\Sigma}_t)$ denotes a multivariate Gaussian distribution evaluated at $\mathbf{x}_t$ defined by mean $\boldsymbol{\mu} \in \mathbb{R}^{H \times W}$ and covariance $\boldsymbol{\Sigma} \in \mathbb{R}^{H \times W}$, where $\boldsymbol{\mu}_t = \sqrt{1-\beta_t}\mathbf{x}_{t-1}$ and $\boldsymbol{\Sigma}_t = \beta_t\mathbf{I}$, with $\mathbf{I} \in \mathbb{R}^{H \times W}$ indicating the identity matrix.

The joint probability of the process that guides the initial sample $\mathbf{x}_0$ to the final sample $\mathbf{x}_T$ is defined in Equation 2.2. This formulation highlights the Markovian characteristic of the process, as each sample $\mathbf{x}_t$ depends only on the immediately preceding sample $\mathbf{x}_{t-1}$ (SOHL-DICKSTEIN et al., 2015).

$$q(\mathbf{x}_{1:T}|\mathbf{x}_0) = \prod_{t=1}^{T} q(\mathbf{x}_t|\mathbf{x}_{t-1}), \tag{2.2}$$

where the trajectory stated by $q(\mathbf{x}_{1:T})$ indicates that the process is repeatedly applied from timestep 1 to T. Once the larger the total number of timesteps T, the greater the computational power required to compute $q(\mathbf{x}_t|\mathbf{x}_{t-1})$. To solve this issue, Ho, Jain e Abbeel (2020) proposed a reparameterization trick that allows a tractable and closed-form of sampling strategy at any given timestep t. Let's assume $\alpha_t = 1 - \beta_t$, then we can use a recursive manner to express $\mathbf{x}_t$ directly in terms of $\mathbf{x}_0$

$$\begin{aligned}
\mathbf{x}_t &= \sqrt{1-\beta_t}\mathbf{x}_{t-1} + \sqrt{\beta_t}\boldsymbol{\epsilon}_{t-1} \\
&= \sqrt{\alpha_t}\mathbf{x}_{t-1} + \sqrt{1-\alpha_t}\boldsymbol{\epsilon}_{t-1} \\
&= \sqrt{\alpha_t}\left(\sqrt{\alpha_{t-1}}\mathbf{x}_{t-2} + \sqrt{1-\alpha_{t-1}}\boldsymbol{\epsilon}_{t-2}\right) + \sqrt{1-\alpha_t}\boldsymbol{\epsilon}_{t-1} \\
&= \sqrt{\alpha_t\alpha_{t-1}}\mathbf{x}_{t-2} + \underbrace{\sqrt{\alpha_t(1-\alpha_{t-1})}\boldsymbol{\epsilon}_{t-2} + \sqrt{1-\alpha_t}\boldsymbol{\epsilon}_{t-1}}_{\text{sum of independent Gaussian}} \\
&= \sqrt{\alpha_t\alpha_{t-1}}\mathbf{x}_{t-2} + \sqrt{1-\alpha_t\alpha_{t-1}}\,\bar{\boldsymbol{\epsilon}}_{t-1} \\
&= \cdots \\
&= \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\,\boldsymbol{\epsilon}
\end{aligned} \tag{2.3}$$

where $\bar{\alpha}_t = \prod_{i=0}^{t}\alpha_i$ and $\boldsymbol{\epsilon}_0, \boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_{t-1} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ are independent Gaussian noise terms. Here, $\mathbf{0} \in \mathbb{R}^{H\times W}$ is a zero matrix, and $\mathbf{I} \in \mathbb{R}^{H\times W}$ is the identity matrix. The merged noise term $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, due to the property that a linear combination of independent Gaussian variables is still Gaussian. By using Equation 2.3, we can define the diffusion kernel of the forward process as shown in Equation 2.4 (HO; JAIN; ABBEEL, 2020)

$$q(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1-\bar{\alpha}_t)\mathbf{I}) \tag{2.4}$$

Finally, Equation 2.4 allows us to generate a sample $\mathbf{x}_t$ directly from an input sample $\mathbf{x}_0$ (HO; JAIN; ABBEEL, 2020). This procss is described in Equation 2.5

$$q(\mathbf{x}_t) = \int q(\mathbf{x}_0, \mathbf{x}_t)d\mathbf{x}_0 = \int q(\mathbf{x}_0)q(\mathbf{x}_t|\mathbf{x}_0)d\mathbf{x}_0, \tag{2.5}$$

where $q(\mathbf{x}_t)$ is the distribution of the intermediate diffused data, $q(\mathbf{x}_0, \mathbf{x}_t)$ is the joint probability, and $q(\mathbf{x}_t|\mathbf{x}_0)$ is the diffusion kernel. To sample $\mathbf{x}_t \sim q(\mathbf{x}_t)$, we first sample $\mathbf{x}_0 \sim q(\mathbf{x}_0)$ and then sample $\mathbf{x}_t \sim q(\mathbf{x}_t|\mathbf{x}_0)$. Figure 6 illustrates how the distributions behaves during this process.

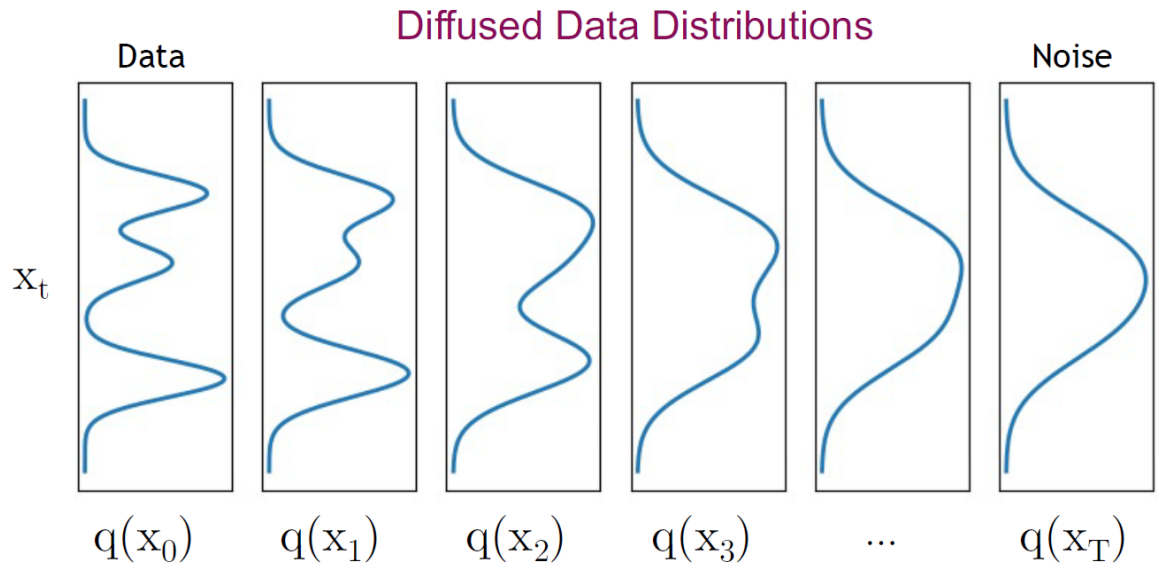


Figure 6 – Illustration of the forward process. Starting from the data distribution $q(\mathbf{x}_0)$, each subsequent step adds more noise, resulting in progressively noisier distributions $q(\mathbf{x}_1), \ldots, q(\mathbf{x}_T)$. As t increases, the distribution gradually approaches an isotropic Gaussian. (**Source:** Kreis, Gao e Vahdat (2022)).

2.1.2 Reverse diffusion process

Ideally, the generative process of diffusion would consist of initially sampling $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and then iteratively computing $\mathbf{x}_{t-1} \sim q(\mathbf{x}_{t-1}|\mathbf{x}_t)$ in a step-wise manner until we recovered a sample $\mathbf{x}_0$ from the original data distribution. Figure 7 illustrates this generative process.

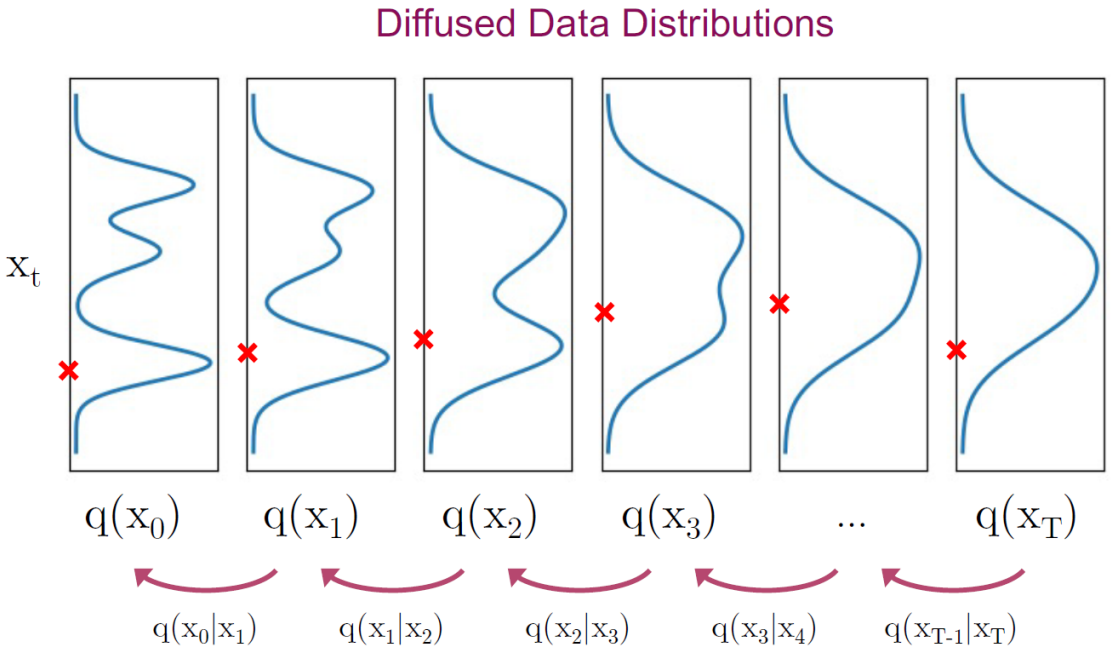


Figure 7 – Illustration of the ideal reverse process. Starting from pure Gaussian noise $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, the model iteratively samples from reverse conditionals $q(\mathbf{x}_{t-1}|\mathbf{x}_t)$, gradually denoising the latent variables to approach a sample $\mathbf{x}_0$ from the data distribution $q(\mathbf{x}_0)$. (**Source:** Kreis, Gao e Vahdat (2022)).

The key challenge lies in modeling the reverse transitions $q(\mathbf{x}_{t-1}|\mathbf{x}_t)$. From the forward process, each latent variable $\mathbf{x}_t$ is conditionally generated from the previous step, according to $q(\mathbf{x}_t|\mathbf{x}_{t-1})$. Now, when generating data, we aim to reverse this process. For that, we require the posterior distribution $q(\mathbf{x}_{t-1}|\mathbf{x}_t)$, which tells us how to sample a cleaner version $\mathbf{x}_{t-1}$ given a noisier one $\mathbf{x}_t$ given a noisier one (HO; JAIN; ABBEEL, 2020). Using Bayes' theorem we can model this process through the reverse conditional distribution expressed in Equation 2.6

$$q(\mathbf{x}_{t-1}|\mathbf{x}_t) = \frac{\overbrace{q(\mathbf{x}_t|\mathbf{x}_{t-1})}^{\text{forward kernel}} \cdot \overbrace{q(\mathbf{x}_{t-1})}^{\text{marginal}}}{\underbrace{q(\mathbf{x}_t)}_{\text{marginal}}} \tag{2.6}$$

Although theoretically defined, this expression is intractable in practice because both $q(\mathbf{x}_{t-1})$ and $q(\mathbf{x}_t)$ are marginal distributions that require integrating on the unknown distribution $q(\mathbf{x}_0)$. This can be written as (HO; JAIN; ABBEEL, 2020)

$$q(\mathbf{x}_{t-1}) = \int q(\mathbf{x}_{t-1}|\mathbf{x}_0)q(\mathbf{x}_0)d\mathbf{x}_0$$

$$q(\mathbf{x}_t) = \int q(\mathbf{x}_t|\mathbf{x}_0)q(\mathbf{x}_0)d\mathbf{x}_0$$

Here the notation $q(\mathbf{x}_0)$ refers to the prior data distribution, and the conditionals $q(\mathbf{x}_{t-1}|\mathbf{x}_0)$ and $q(\mathbf{x}_t|\mathbf{x}_0)$ are known Gaussian transitions. What is fundamentally difficult to compute are the reverse conditionals $q(\mathbf{x}_{t-1}|\mathbf{x}_t)$ due to the unknown nature of $q(\mathbf{x}_0)$ (LUO, 2022; STRüMKE; LANGSETH, 2023; GALLON; JENTZEN; WURSTEMBERGER, 2024). This intractability motivated the use of learned neural network approximation $p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)$ to approximate $q(\mathbf{x}_{t-1}|\mathbf{x}_t)$

Since the reverse process has the same distributional form as the forward process (FELLER, 1949), which is an isotropic Gaussian distribution, then the goal of the reverse process is to learn the parameters of the Gaussian transitions that revert the forward process. These parameters, specifically the mean and variance, can be predicted by a deep neural network (SOHL-DICKSTEIN et al., 2015).

Objectively, we use a deep neural network p_θ parameterized by θ to approximate the reverse transition distribution $q(\mathbf{x}_{t-1}|\mathbf{x}_t)$. This distribution is modeled as a Gaussian whose parameters depend on the noisy input $\mathbf{x}_t$ and the corresponding timestep t:

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t)), \tag{2.7}$$

where $\boldsymbol{\mu}_\theta$ and $\boldsymbol{\Sigma}_\theta$ denote the predicted mean and covariance of the reverse Gaussian transition. By conditioning the model on both the timestep t and the noisy sample $\mathbf{x}_t$, the network learns to denoise different levels of noise added in the forward process. Figure 8 illustrates this process.

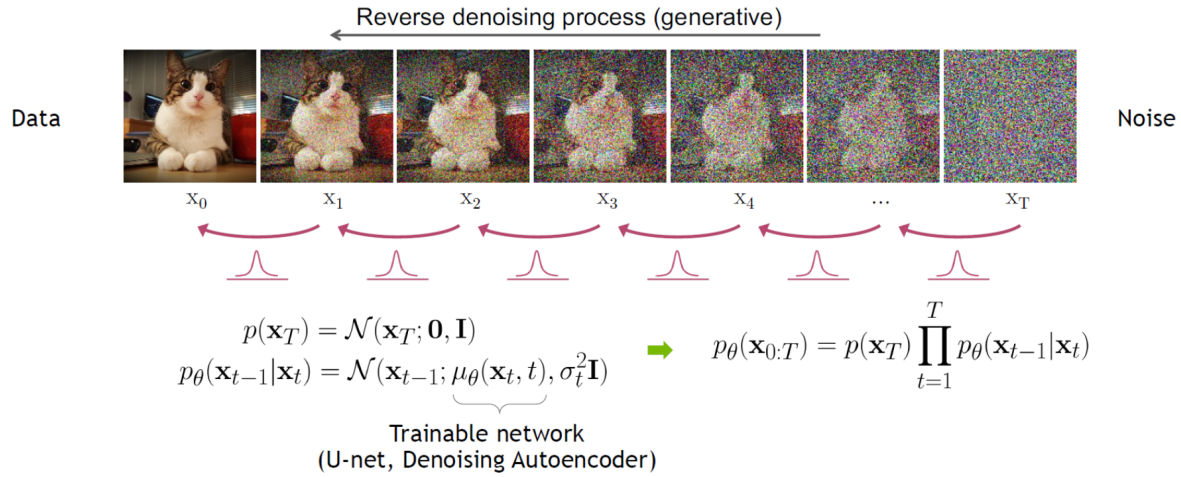


Figure 8 – Illustration of the generative (reverse) diffusion process. Starting from a noise sample $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, the model iteratively reconstructs a data-like sample $\hat{\mathbf{x}}_0$ by applying a trained neural network. (**Source:** Kreis, Gao e Vahdat (2022)).

Just like in the forward diffusion, the reverse process can also be modeled as a Markov chain, trained to recover $\hat{\mathbf{x}}_0$ from pure noise $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ (HO; JAIN; ABBEEL, 2020). Consequently, the full generative model is expressed as

$$p_\theta(\mathbf{x}_{0:T}) = p(\mathbf{x}_T) \prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t), \tag{2.8}$$

where the $p_\theta(\mathbf{x}_{0:T})$ trajectory represents the joint distribution over all latent variables along the reverse Markov chain.

2.2 TRAINING OBJECTIVE

Like most generative models, the training objective of diffusion models is to learn a model distribution $p_\theta(\mathbf{x}_0)$ that approximates the true data distribution $q(\mathbf{x}_0)$, which is achieved by minimizing the negative log-likelihood via a variational lower bound (SOHL-DICKSTEIN et al., 2015)

$$\begin{aligned} -\log p_\theta(\mathbf{x}_0) &\leq -\log p_\theta(\mathbf{x}_0) + D_{\mathrm{KL}}\left(q(\mathbf{x}_{1:T}|\mathbf{x}_0) \| p_\theta(\mathbf{x}_{1:T}|\mathbf{x}_0)\right) \\ &= -\log p_\theta(\mathbf{x}_0) + \mathbb{E}_{\mathbf{x}_{1:T}\sim q(\mathbf{x}_{1:T}|\mathbf{x}_0)}\left[\log \frac{q(\mathbf{x}_{1:T}|\mathbf{x}_0)}{p_\theta(\mathbf{x}_{0:T}) / p_\theta(\mathbf{x}_0)}\right] \\ &= -\log p_\theta(\mathbf{x}_0) + \mathbb{E}_q\left[\log \frac{q(\mathbf{x}_{1:T}|\mathbf{x}_0)}{p_\theta(\mathbf{x}_{0:T})} + \log p_\theta(\mathbf{x}_0)\right] \\ &= \mathbb{E}_q\left[\log \frac{q(\mathbf{x}_{1:T}|\mathbf{x}_0)}{p_\theta(\mathbf{x}_{0:T})}\right] \end{aligned} \tag{2.9}$$

Equation 2.9 yields the following VLB objetive

$$L_{\mathrm{VLB}} = \mathbb{E}_{q(\mathbf{x}_0)q(\mathbf{x}_{1:T}|\mathbf{x}_0)}\left[-\log \frac{p_\theta(\mathbf{x}_{0:T})}{q(\mathbf{x}_{1:T}|\mathbf{x}_0)}\right] \geq \mathbb{E}_{q(\mathbf{x}_0)}\left[-\log p_\theta(\mathbf{x}_0)\right],$$

2.2.1 Decomposition of the VLB

Using the Markov property of both processes, we decompose L_{VLB} into a combination of several KL-divergence terms (HO; JAIN; ABBEEL, 2020)

$$
\begin{aligned}
L_{\text{VLB}} &= \mathbb{E}_{q(\mathbf{x}_{0,T})}\left[\log \frac{q\left(\mathbf{x}_{1:T}|\mathbf{x}_0\right)}{p_\theta\left(\mathbf{x}_{0:T}\right)}\right] \\
&= \mathbb{E}_q\left[\log \frac{\prod_{t=1}^{T} q\left(\mathbf{x}_t|\mathbf{x}_{t-1}\right)}{p_\theta\left(\mathbf{x}_T\right)\prod_{t=1}^{T} p_\theta\left(\mathbf{x}_{t-1}|\mathbf{x}_t\right)}\right] \\
&= \mathbb{E}_q\left[-\log p_\theta\left(\mathbf{x}_T\right) + \sum_{t=1}^{T}\log \frac{q\left(\mathbf{x}_t|\mathbf{x}_{t-1}\right)}{p_\theta\left(\mathbf{x}_{t-1}|\mathbf{x}_t\right)}\right] \\
&= \mathbb{E}_q\left[-\log p_\theta\left(\mathbf{x}_T\right) + \sum_{t=2}^{T}\log \frac{q\left(\mathbf{x}_t|\mathbf{x}_{t-1}\right)}{p_\theta\left(\mathbf{x}_{t-1}|\mathbf{x}_t\right)} + \log \frac{q\left(\mathbf{x}_1|\mathbf{x}_0\right)}{p_\theta\left(\mathbf{x}_0|\mathbf{x}_1\right)}\right] \\
&= \mathbb{E}_q\left[-\log p_\theta\left(\mathbf{x}_T\right) + \sum_{t=2}^{T}\log \left(\frac{q\left(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0\right)}{p_\theta\left(\mathbf{x}_{t-1}|\mathbf{x}_t\right)}\cdot\frac{q\left(\mathbf{x}_t|\mathbf{x}_0\right)}{q\left(\mathbf{x}_{t-1}|\mathbf{x}_0\right)}\right) + \log \frac{q\left(\mathbf{x}_1|\mathbf{x}_0\right)}{p_\theta\left(\mathbf{x}_0|\mathbf{x}_1\right)}\right] \\
&= \mathbb{E}_q\left[-\log p_\theta\left(\mathbf{x}_T\right) + \sum_{t=2}^{T}\log \frac{q\left(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0\right)}{p_\theta\left(\mathbf{x}_{t-1}|\mathbf{x}_t\right)} + \sum_{t=2}^{T}\log \frac{q\left(\mathbf{x}_t|\mathbf{x}_0\right)}{q\left(\mathbf{x}_{t-1}|\mathbf{x}_0\right)} + \log \frac{q\left(\mathbf{x}_1|\mathbf{x}_0\right)}{p_\theta\left(\mathbf{x}_0|\mathbf{x}_1\right)}\right] \\
&= \mathbb{E}_q\left[-\log p_\theta\left(\mathbf{x}_T\right) + \sum_{t=2}^{T}\log \frac{q\left(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0\right)}{p_\theta\left(\mathbf{x}_{t-1}|\mathbf{x}_t\right)} + \log \frac{q\left(\mathbf{x}_T|\mathbf{x}_0\right)}{q\left(\mathbf{x}_1|\mathbf{x}_0\right)} + \log \frac{q\left(\mathbf{x}_1|\mathbf{x}_0\right)}{p_\theta\left(\mathbf{x}_0|\mathbf{x}_1\right)}\right] \\
&= \mathbb{E}_q\left[\log \frac{q\left(\mathbf{x}_T|\mathbf{x}_0\right)}{p_\theta\left(\mathbf{x}_T\right)} + \sum_{t=2}^{T}\log \frac{q\left(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0\right)}{p_\theta\left(\mathbf{x}_{t-1}|\mathbf{x}_t\right)} - \log p_\theta\left(\mathbf{x}_0|\mathbf{x}_1\right)\right] \\
&= \mathbb{E}_q\left[\underbrace{D_{KL}(q(\mathbf{x}_T|\mathbf{x}_0)||p(\mathbf{x}_T))}_{L_T} + \sum_{t>1}\underbrace{D_{KL}(q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0)||p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t))}_{L_{t-1}}\underbrace{-\log p_\theta(\mathbf{x}_0|\mathbf{x}_1)}_{L_0}\right]
\end{aligned}
\tag{2.10}
$$

The term L_T shows how close $\mathbf{x}_T$ is to the standard Gaussian. Since the approximate posterior q has no learnable parameters due to the fixed variance schedule, this term is constant during training and can be ignored. L_{t-1}, or simply L_t, expresses the difference between the desired denoising steps $p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)$ and the approximate ones $q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0)$. The term L_0 is similar to the reconstruction term of a variational autoencoder (KINGMA et al., 2016; SALIMANS et al., 2017).

2.2.2 Parameterization of L_t for training loss

In practice, Ho, Jain e Abbeel (2020) derived a tractable reverse conditional probability $q(\mathbf{x}_{t-1}|\mathbf{x}_t)$ by additionally conditioning it on $\mathbf{x}_0$. The new formulation of this process is written

in Equation 2.11

$$q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_{t-1}; \tilde{\boldsymbol{\mu}}_t(\mathbf{x}_t, \mathbf{x}_0), \tilde{\beta}_t \mathbf{I}), \tag{2.11}$$

Since $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{i=0}^{t} \alpha_i$, the variance can be defined as follows

$$\tilde{\beta}_t = 1 \Big/ \left(\frac{\alpha_t}{\beta_t} + \frac{1}{1-\bar{\alpha}_t}\right) = 1 \Big/ \left(\frac{\alpha_t - \bar{\alpha}_t + \beta_t}{\beta_t\,(1-\bar{\alpha}_{t-1})}\right) = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t} \cdot \beta_t$$

Following the standard Gaussian density function, the mean can be parameterized as

$$\begin{aligned}
\tilde{\boldsymbol{\mu}}_t(\mathbf{x}_t, \mathbf{x}_0) &= \left(\frac{\sqrt{\alpha_t}}{\beta_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}}\mathbf{x}_0\right) \Big/ \left(\frac{\alpha_t}{\beta_t} + \frac{1}{1-\bar{\alpha}_{t-1}}\right) \\
&= \left(\frac{\sqrt{\alpha_t}}{\beta_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}}{1-\bar{\alpha}_{t-1}}\mathbf{x}_0\right) \cdot \left(\frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t} \cdot \beta_t\right) \\
&= \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\mathbf{x}_0 + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t
\end{aligned} \tag{2.12}$$

From the reparameterization trick expressed in Equation 2.3, we can write $\mathbf{x}_0$ as

$$\mathbf{x}_0 = \frac{1}{\sqrt{\alpha_t}}(\mathbf{x}_t - \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}) \tag{2.13}$$

Finally, by replacing Equation 2.13 in Equation 2.12, Ho, Jain e Abbeel (2020) showed the compact form of the estimated mean in Equation 2.15

$$\tilde{\boldsymbol{\mu}}_t(\mathbf{x}_t, \mathbf{x}_0) = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t - \frac{1-\bar{\alpha}_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}\right) \tag{2.14}$$

From Equation 2.7, we know we need to learn a neural network to approximate the conditioned probability distributions in the reverse diffusion process

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t)),$$

which means we would like to train $\boldsymbol{\mu}_\theta$ to predict $\tilde{\boldsymbol{\mu}}_t$. Since $\mathbf{x}_t$ is available as input at training time, we can reparameterize the Gaussian noise term instead to make it predict $\boldsymbol{\epsilon}_t$ from the input $\mathbf{x}_t$ at timestep t

$$\boldsymbol{\mu}_\theta(\mathbf{x}_t, \mathbf{x}_0) = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t - \frac{1-\bar{\alpha}_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)\right), \tag{2.15}$$

where $\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)$ is the noise predicted by the network as timestep t.

From Equation 2.10, the loss term L_t can be parameterized to minimize the difference from $\tilde{\boldsymbol{\mu}}_t$ (HO; JAIN; ABBEEL, 2020)

$$L_t = D_{\mathrm{KL}}\left(q\left(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0\right) \| p_\theta\left(\mathbf{x}_{t-1}|\mathbf{x}_t\right)\right) = \mathbb{E}_q\left[\frac{1}{2\sigma_t^2}\left\|\tilde{\mu}_t\left(\mathbf{x}_t, \mathbf{x}_0\right) - \mu_\theta\left(\mathbf{x}_t, t\right)\right\|^2\right]$$

Ho, Jain e Abbeel (2020) demonstrated that maximizing the log-likelihood can be reduced to simply learning the denoising steps L_t

$$
\begin{aligned}
\boldsymbol{L_t} &= \mathbb{E}_{\mathbf{x_0},\boldsymbol{\epsilon}}\left[\frac{1}{2\left\|\boldsymbol{\Sigma_\theta}\left(\mathbf{x}_t,\boldsymbol{t}\right)\right\|_{\mathbf{2}}^2}\left\|\tilde{\mu}_t\left(\mathrm{x}_t,\mathrm{x}_0\right)-\mu_\theta\left(\mathrm{x}_t,t\right)\right\|^2\right] \\
&= \mathbb{E}_{\mathbf{x_0},\boldsymbol{\epsilon}}\left[\frac{1}{2\left\|\boldsymbol{\Sigma_\theta}\right\|_2^2}\left\|\frac{1}{\sqrt{\alpha_t}}\left(\mathrm{x}_t-\frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}\right)-\frac{1}{\sqrt{\alpha_t}}\left(\mathrm{x}_t-\frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta\left(\mathrm{x}_t,t\right)\right)\right\|^2\right] \\
&= \mathbb{E}_{\mathbf{x_0},\boldsymbol{\epsilon}}\left[\frac{\left(1-\boldsymbol{\alpha_t}\right)^2}{2\boldsymbol{\alpha}_t\left(1-\bar{\alpha}_{\boldsymbol{t}}\right)\left\|\boldsymbol{\Sigma_\theta}\right\|_{\mathbf{2}}^2}\left\|\boldsymbol{\epsilon}-\boldsymbol{\epsilon_\theta}\left(\mathbf{x}_{\boldsymbol{t}},\boldsymbol{t}\right)\right\|^2\right] \\
&= \mathbb{E}_{\mathbf{x_0},\boldsymbol{\epsilon}}\left[\underbrace{\frac{\left(1-\boldsymbol{\alpha_t}\right)^{\mathbf{2}}}{2\boldsymbol{\alpha_t}\left(1-\bar{\alpha}_{\boldsymbol{t}}\right)\left\|\boldsymbol{\Sigma_\theta}\right\|_{\mathbf{2}}^2}}_{\text{weighting factor}}\left\|\boldsymbol{\epsilon_t}-\boldsymbol{\epsilon_\theta}\left(\underbrace{\sqrt{\bar{\alpha}_{\boldsymbol{t}}}\mathbf{x_0}+\sqrt{\mathbf{1}-\bar{\alpha}_{\boldsymbol{t}}}\boldsymbol{\epsilon}}_{\mathbf{x}_t},\boldsymbol{t}\right)\right\|^2\right]
\end{aligned}
\tag{2.16}
$$

Ho, Jain e Abbeel (2020) showed empirically that better results were achieved with a simplified objective that ignores the weighting term

$$
L_t^{simple} = \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}\left[||\boldsymbol{\epsilon}-\boldsymbol{\epsilon}_\theta(\mathbf{x}_t,t)||^2\right] \tag{2.17}
$$

In practical terms, this corresponds to the mean squared error between the actual noise at time t, $\boldsymbol{\epsilon}_t$, and the predicted noise by the network at time t, $\boldsymbol{\epsilon}_\theta(\mathbf{x}_t,t)$, given a sample $\mathbf{x}_t$.

Algorithm 1 (HO; JAIN; ABBEEL, 2020) summarizes the training process for a standard diffusion model. The main objective is to train the network to predict the noise added to an image at a given timestep. At each iteration, a clean image $\mathbf{x}_0$ is sampled from the data distribution and a timestep t is chosen randomly. Then, synthetic Gaussian noise is sampled $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0},\mathbf{I})$ and used to generate a noisy version of the input image according to Equation 2.3. The network parameters are then optimized via gradient descent to minimize the squared error between the actual noise $\boldsymbol{\epsilon}$ and the prediction $\boldsymbol{\epsilon}_\theta(\mathbf{x}_t,t)$.

Algorithm 1 Training

1: **Input:** data distribution $q(\mathbf{x}_0)$, total timesteps T
2: **Output** trained network $\boldsymbol{\epsilon}_\theta$
3: **repeat**
4: $\quad \mathbf{x}_0 \sim q(\mathbf{x}_0)$
5: $\quad t \sim$ Uniform$(\{1,\ldots,T\})$
6: $\quad \boldsymbol{\epsilon} \sim \mathcal{N}(0,\mathbf{I})$
7: $\quad$ Take gradient descent step on
8: $\quad\quad \nabla_\theta\|\boldsymbol{\epsilon}-\boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0+\sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon},t)\|^2$
9: **until** converged

During the inference phase, Algorithm 2 (HO; JAIN; ABBEEL, 2020) is used to generate new image samples. In short, the model uses the predictions of the trained network to progressively

reverse the noise addition process, starting from a pure noise sample $\mathbf{x}_T \sim \mathcal{N}(0, \mathbf{I})$. Going through each timestep, from T to 1, the network uses the noise estimate present in the current image $\mathbf{x}_t$ to reconstruct the less noisy image $\mathbf{x}_{t-1}$, according to Equation 2.15. The term $\mathbf{z}$ is an additional noise term used to control variability, and $\boldsymbol{\sigma}_t^2 = \tilde{\beta}_t$. At the end of the process, a new image sample $\mathbf{x}_0$ is obtained.

Algorithm 2 Sampling

1: $\mathbf{x}_T \sim \mathcal{N}(0, \mathbf{I})$
2: **for** $t = T, \ldots, 1$ **do**
3: $\quad \mathbf{z} \sim \mathcal{N}(0, \mathbf{I})$ if $t > 1$, else $\mathbf{z} = \mathbf{0}$
4: $\quad \mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\alpha_t}} \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t) \right) + \boldsymbol{\Sigma}_t \mathbf{z}$
5: **end for**
6: **return** $\mathbf{x}_0$

2.3 MODEL ARCHITECTURE

The original implementation of DDPMs (HO; JAIN; ABBEEL, 2020) employs a U-Net model (RONNEBERGER; FISCHER; BROX, 2015), with ResNet blocks (HE et al., 2016), group normalization (WU; HE, 2018) as well as self-attention blocks (VASWANI et al., 2017). Each diffusion timesteps t is specified to the network by adding a sinusoidal position embedding (VASWANI et al., 2017) into each residual block.

To provide the model a sense of progression through time, the timestep t is mapped to a high-dimension vector $\lambda(t) \in \mathbb{R}^d$ using a sinusoidal encoding defined as

$$\lambda(t)_{2i} = \sin\left(\frac{t}{10000^{2i/d}}\right)$$
$$\lambda(t)_{2i+1} = \cos\left(\frac{t}{10000^{2i/d}}\right),$$

where d is the embedding dimension (VASWANI et al., 2017). After this, the embedding $\lambda(t)$ is passed to a sequence of fully-connected layers producing a learned vector representation that is added into the blocks of the U-Net. This temporal embedding enables the network to adjust its behavior depending on how much noise is present in the input image $\mathbf{x}_t$ (DHARIWAL; NICHOL, 2021). Figure 9 illustrates this process.

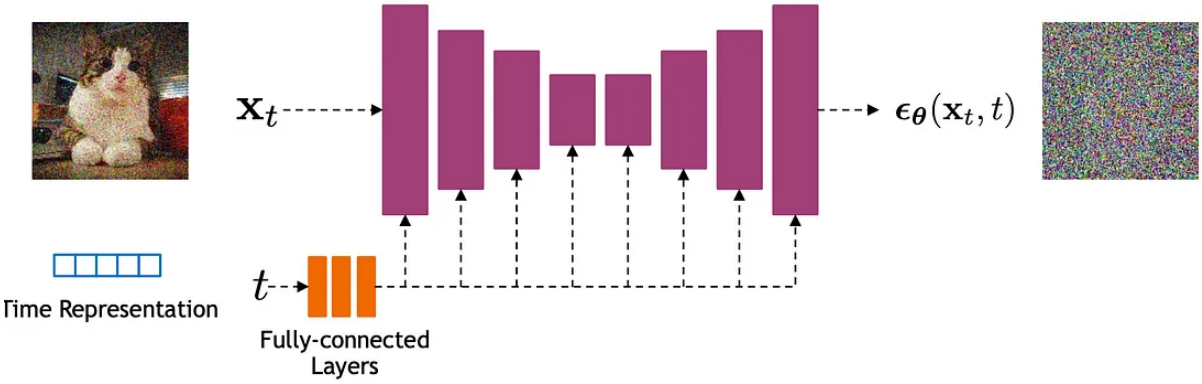


Figure 9 – Illustration of the denoising neural network architecture used in DDPMs. At each training step t, a noisy image $\mathbf{x}_t$ is input to a U-Net-based model. The timestep t is encoded via a sinusoidal position embedding and processed through fully-connected layers before being injected into the blocks of the U-Net. The model is trained to predict the noise $\epsilon_\theta(\mathbf{x}_t, t)$ added to the original image. (**Source:** Kreis, Gao e Vahdat (2022)).

2.4 NOISE SCHEDULE

Each diffusion step is modeled by a variance schedule $\{\beta_t\}_{t=1}^{T}$ that dictates the diffusion rate, that is how fast the data distribution will be converted into the prior distribution, which is the Gaussian in the standard formulation. To achieve this, the noise addition must follow a specific pace, controlled by a chosen schedule. The variance parameter β_t can be fixed to a constant or follow a specific schedule over the T timesteps. Ho, Jain e Abbeel (2020) set $T = 1000$ and used a linear schedule that increases from $\beta_1 = 10^{-4}$ to $\beta_T = 0.02$, while Nichol e Dhariwal (2021) employed a cosine schedule. Guo et al. (2025) examined several types of noise schedule, such as the Fibonacci schedule, which is derived from the first T terms of the Fibonacci sequence (CHEN et al., 2020); the Laplace schedule, which is based on the Laplace distribution (HANG et al., 2024); and even a learned noise schedule instead of a fixed one, modeled by a monotonic neural network (SAHOO et al., 2024). Figure 10 shows the $\bar{\alpha}_t$ progression for the linear and cosine schedules.

The work of Chen (2023) expanded the study on the noise addition aspect of diffusion models by examining the performance of different choices of noise schedule conditioned on different image sizes. Figure 11 highlights the importance of choosing an appropriate noise schedule. As the image resolution increases, the denoising task becomes simpler because the redundancy of information in data increases, making it easier to recover the original signal. Some practical observations demonstrate that, as image size increases, the optimal noise scheduling tends to shift towards a noisier one due to increased redundancy in pixels (GU et al., 2022; HOOGEBOOM; HEEK; SALIMANS, 2023).

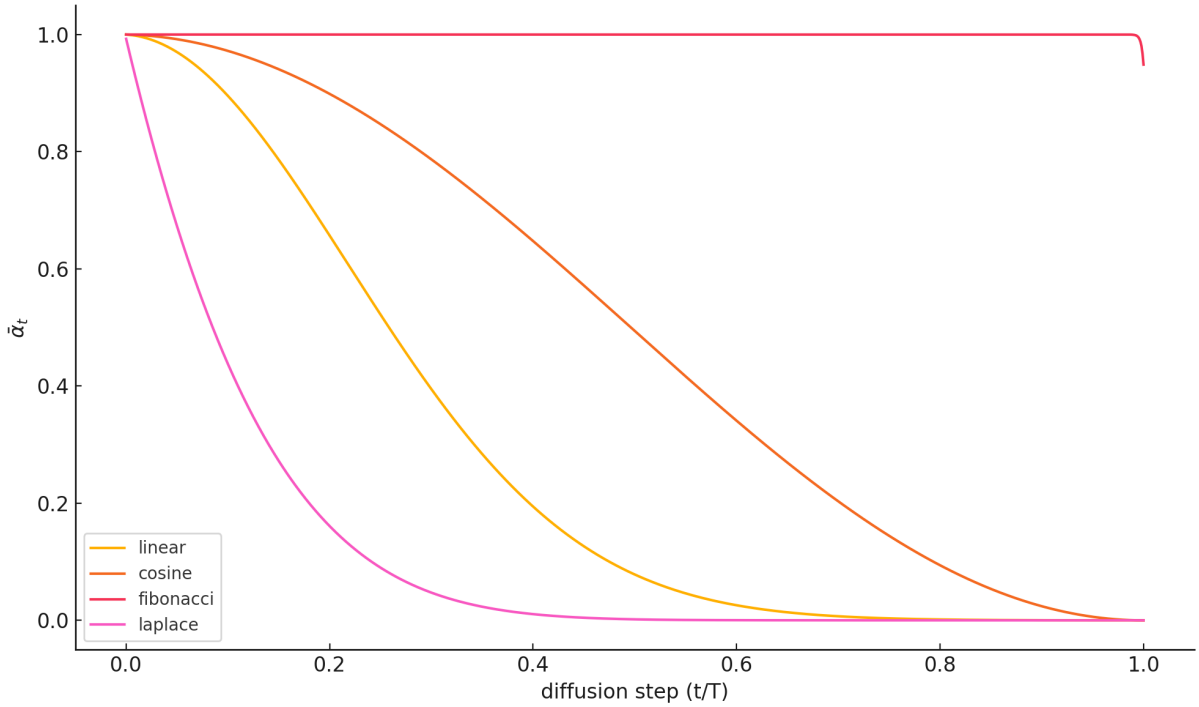
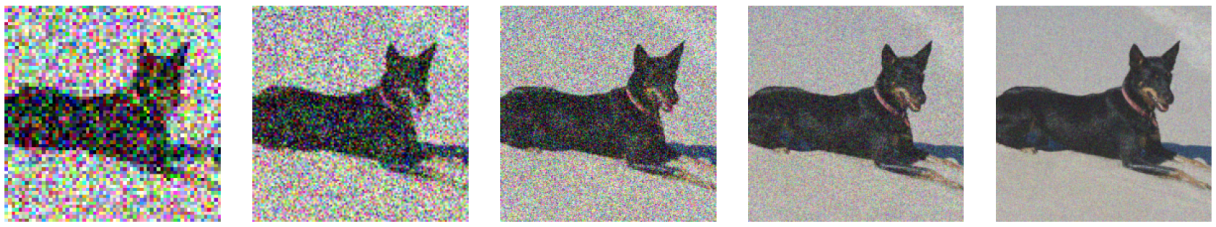


Figure 10 – $\bar{\alpha}_t$ throughout the diffusion steps for different schedules. The yellow curve represents the linear schedule, where the steadily increase of noise variance over time results in a faster decrease of α_t. The orange curve shows the cosine schedule, which preserves more information from the original image in the first steps. The red curve is the Fibonacci schedule, which produces slower nonlinear growth at first, with noise intensifying in the final steps. The Laplace schedule, represented by the pink curve, concentrates noise in the initial and final steps, with less noise in the center. (**Source:** Adapted from Nichol e Dhariwal (2021)).



(a) 64×64 (b) 128×128 (c) 256×256 (d) 512×512 (e) 1024×1024

Figure 11 – Example of noised images corrupted by the same noise level. Higher-resolution images tend to exhibit more overall local redundancy between neighboring pixels, which makes the effect of noise less noticeable. That means less information is destroyed with the same level of noise. (**Source:** Chen (2023)).

Nichol e Dhariwal (2021) also explores the model efficiency related to the choice of total noise steps T. Just as the image size, the value of T also correlates to the choice of noise schedule. Their work also shows that the linear noise schedule used in Ho, Jain e Abbeel (2020) worked well for high-resolution images but was suboptimal for images of resolution 64×64 and 32×32, because The linear schedule's forward noising process becomes too noisy at the end, which reduces its contributions to sample quality. Finally, the authors indicate that the noise schedule should be adapted based on the choice of T and the image resolution

(NICHOL; DHARIWAL, 2021). Figure 12 demonstrates a visual comparison between the noise addition in the linear and cosine schedules.

Figure 12 – Example of latent samples generated with linear (top) and cosine (bottom) noise schedules respectively at linearly spaced values of t from 0 to T. In the linear schedule, the last latent states are almost entirely pure noise, while the cosine schedule adds noise more gradually, preserving the overall visual structure for more steps. (**Source:** Nichol e Dhariwal (2021)).

3 RELATED WORKS

This chapter reviews some of the most relevant studies related to LDCT denoising and reconstruction techniques. It first presents a brief overview of these methods in Section 3.1. Section 3.2, explores conditioning mechanisms in diffusion models. Finally, Sections 3.3 and 3.4 detail image-conditioned diffusion methods and their recent advancements, which represent the current state of the art in generative reconstruction frameworks.

3.1 LDCT IMAGE RECONSTRUCTION AND DENOISING TECHNIQUES

Low-dose computed tomography has became a central research topic in medical imaging due to its potential to minimize radiation exposure while maintaining diagnostically reliable image quality. The main challenge in LDCT imaging is the presence of high levels of noise caused by the reduced X-ray flux, which manifests as artifacts that degrade image interpretability. Over the years, several LDCT image reconstruction and denoising techniques have been proposed to mitigate these effects, ranging from traditional model-based algorithms to modern deep learning methods.

Early approaches to LDCT reconstruction were grounded in analytical and iterative formulations (MURPHY et al., 2025). The Filtered Back Projection (FBP) (PELT; BATENBURG, 2013) method served as the foundation of CT reconstruction and was often used alongside machine learnig based reconstruction techniques to generate CT images. To improve upon FBP, researchers developed model-based iterative reconstruction (MBIR) (LIU, 2014) methods, which incorporate statistical priors and noise models into the reconstruction process. MBIR and its variants, such as adaptive statistical iterative reconstruction (ASIR) (HSIEH et al., 2013) and penalized weighted least-squares (PWLS) (THIBAULT et al., 2007) approaches, were able to reduce noise while preserving structural details through explicit regularization terms. However, iterative methods suffer from high computational cost and sensitivity to noise artifacts (QIN et al., 2019).

With the advent of deep learning techniques, data-driven models rapidly became the dominant approach to LDCT denoising. Convolutional Neural Networks were among the first to achieve good performance in suppressing CT noise. For instance, the Residual Encoder-Decoder Convolutional Neural Network (RED-CNN) proposed by Chen et al. (2017) combines convo-

lutional layers with residual learning to restore NDCT images from their LDCT counterparts. Despite their success, CNN-based models tend to oversmooth fine details, compromising texture quality of anatomical structures (WOLTERINK et al., 2017).

To overcome these limitations, generative models were introduced to better capture the distribution of high-quality CT images. Generative Adversarial Networks revolutionized the synthetic data generation paradigm, particularly in imaging applications. Using an adversarial structure, composed of two competing networks, a generator and a discriminator, GANs are able to produce data indistinguishable from real data. The work of Wolterink et al. (2017) demonstrated the ability of GANs to synthesize high-quality images from noisy inputs. Expanding this paradigm, the Wasserstein Generative Adversarial Network (WGAN) (ARJOVSKY; CHINTALA; BOTTOU, 2017) model introduced a perceptual loss to enhance visual realism and preserve texture details, addressing the oversmoothing effect observed in MSE-based methods.

An advanced variant of the GAN is the Self-Attention Convolutional Neural Network (SACNN) proposed by Li et al. (2020), which integrates a self-attention mechanism into the generator architecture. The self-attention mechanism gives the SACNN the ability to capture long-range contextual dependencies across the entire image, essential for detailed image reconstruction, which solidifies it as a robust model for LDCT reconstruction.

In more recent years, diffusion models emerged as a promising class of generative frameworks capable of learning complex data distributions through a more stable training process than the adversarial paradigm of GANs. The following sections explore conditional diffusion models and their extensions to image-guided restoration tasks, as well as the current state of the art in diffusion-based approaches for image denoising.

3.2 CONDITIONAL DIFFUSION MODELS

Conditioning is one of the most important aspects of generative models. In the context of diffusion models, guidance can refer to a textual prompt, a class label or an image (HO et al., 2022). Mathematically, this refers to conditioning a prior distribution $p(\mathbf{x})$ with a condition $\mathbf{y}$, resulting in $p(\mathbf{x}|\mathbf{y})$ (HO et al., 2022). A conditional diffusion model can be modeled as

$$p_\theta(\mathbf{x}_{0:T}|y) = p(\mathbf{x}_T) \prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t, y), \tag{3.1}$$

where the conditioning information $\mathbf{y}$ is incorporated at each timestep t. One of the reasons behind the high sampling quality of diffusion models, particularly in text-to-image generation,

is precisely the fact that the condition is made available to the model at every denoising step (NICHOL et al., 2021; SAHARIA et al., 2022b; ROMBACH et al., 2022; RAMESH et al., 2022).

3.3 IMAGE-CONDITIONED DIFFUSION

Since the popularization of conditional diffusion models, several approaches were developed in the field of inverse problems and deterministic image generation (KADKHODAIE; SIMONCELLI, 2021; KAWAR; VAKSMAN; ELAD, 2021; KAWAR et al., 2022). These methods apply the idea of feeding an input image as a conditional information into the network. This concept has been successfully adapted to many models designed for the problems of deblurring, inpainting, super-resolution, image enhancement and image restoration (SAHARIA et al., 2022a; SAHARIA et al., 2022c; WHANG et al., 2022; ÖZDENIZCI; LEGENSTEIN, 2023; GUO et al., 2023; HOU et al., 2023; WANG et al., 2025).

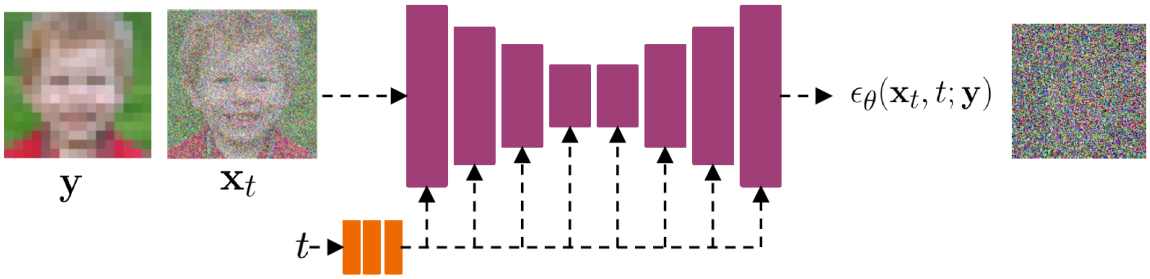


Figure 13 – Illustration of the general image conditioning pipeline applied to DDPMs. Typically, the condition image $\mathbf{y}$ is concatenated to the network along with $\mathbf{x}_t$. (**Source:** Saharia et al. (2022c)).

As explained in Chapter 2, both the training and inference stages of the standard diffusion model pipeline rely heavily on the properties of the Gaussian noise applied during the forward and reverse processes. In an effort to deconstruct this reliance, Bansal et al. (2022) introduced Cold Diffusion, a framework that extends the classical diffusion formulation to arbitrary image transformations. In contrast to traditional noise-based approaches, that rely on the addition of Gaussian noise throughout the forward process, the degradation operator D of Cold Diffusion applies specific image transformations such as blurring or masking, which are fully deterministic. On the other hand, the restoration operator R is trained to invert these transformations using the ℓ_1 norm as loss function, producing a generative behavior. The authors generalize their method to several different tasks, such as deblurring, inpainting, super-resolution and snowification (BANSAL et al., 2022).

One of the main contributions of Cold Diffusion consists in the improved sampling technique shown in Algorithm 3 (BANSAL et al., 2022). The traditional diffusion sampling process, explained in Algorithm 2, adds noise back to the image during iterative denoising in order to improve stability. Similarly, the authors verified that adding a certain level of degradation into the restoration process produces greater stability in regard to the errors yielded by the network (BANSAL et al., 2022; XIE et al., 2023).

Algorithm 3 Improved Sampling for Cold Diffusion

Input: a degraded sample $\mathbf{x}_t$
for $t = T, \ldots, 1$ **do**
    $\hat{\mathbf{x}}_0 = R(\mathbf{x}_t, t)$
    $\mathbf{x}_{t-1} = \mathbf{x}_t - D(\hat{\mathbf{x}}_0, t) + D(\hat{\mathbf{x}}_0, t-1)$
end for
return $\mathbf{x}_0$

3.4 DIFFUSION FOR IMAGE DENOISING

In the context of image denoising, the use of the traditional diffusion-based model usually proves inefficient, as the modeled noise is not necessarily completely consistent with the Gaussian noise applied during the diffusion process (XIE et al., 2023). As a result, the current state-of-the-art methods rely on sophisticated strategies to overcome this issue.

Xu et al. (2024) introduce SWORD, a Stage-by-stage Wavelet Optimization Refinement Diffusion model for sparse-view CT reconstruction. Their method integrates wavelet decomposition into the diffusion framework. Existing diffusion models often focus on the sinogram or image domains, leading to instability during training and convergence to local minimal solutions. Unlike these models, SWORD establishes a unified mathematical model incorporating low-frequency and high-frequency, solved through an optimization procedure. By refining the reconstruction at each stage through wavelet-based priors, the model is able to effectively remove noise artifacts.

Although supervised deep learning techniques have shown promise in tackling the problem of LDCT denoising, they require a set of paired low-dose and normal-dose CT images for training, which can be difficult to acquire in clinical settings. To address this limitation, Liu et al. (2023) introduce Dn-Dp, an unsupervised method for denoising LDCT images using diffusion probabilistic priors. The proposed method uses NDCT images for training, enabling zero-shot LDCT image denoising.

Shi et al. (2024) introduce Resfusion, a general framework for image restoration using denoising diffusion probabilistic models. Unlike diffusion-based methods that rely on degraded images as conditional input, Resfusion incorporates a residual term into the forward process, initiating the reverse process directly from noisy degraded images, rather than pure Gaussian noise itself, enabling more effective restoration and noise removal.

Finally, Gao et al. (2023) introduce CoreDiff, a novel COntextual eRror-modulated gEneralized Diffusion model, specifically designed for LDCT denoising and generalization. Their work aims to reduce the long inference times associated with standard diffusion models by extending the Cold Diffusion model. CoreDiff uses LDCT images instead of random Gaussian noise, employing a "mean-preserving degradation operator" to mimic CT degradation. Furthermore, the authors propose a ContextuaL Error-modulAted Restorarion Network, CLEAR-Net, which uses contextual information to constrain the sampling process and refine timestep embeddings. A one-shot learning framework is also introduced, which enables the rapid adaptation of the trained model to unseen dose levels.

4 CONDITIONAL DIFFUSION FOR LDCT DENOISING

In this chapter, we present our diffusion-based framework developed specifically for LDCT image denoising. The method builds upon existing strategies that incorporates the conditioning noisy image directly into the forward process, an approach that better reflects the characteristics of real-world noise and eliminates the need to assume a specific noise distribution.

While such image-conditioned formulation has been previously explored, our contribution lies in extending its noise schedule aspect through a novel modulation mechanism. Specifically, we introduce a parameter γ that controls the schedule by which noise is distributed throughout the diffusion process.

The remainder of this chapter introduced the full training and inference algorithms in Sections 4.1 and 4.3, respectively, as well as an in-depth analysis of the proposed modifications to the forward and reverse processes, presented in Section 4.2.

4.1 TRAINING OVERVIEW

Our training process follows the standard diffusion approach presented in Section 2.2, which consists of learning a denoising function employing a parametrized U-Net model $\mathbf{U}_\theta$. The network takes a noisy observation $\mathbf{x}_t$, generated via an interpolation method described in Section 4.2.1, and a timestep t, which is embedded into $\mathbf{U}_\theta$ using sinusoidal positional encoding, as inputs and predicts the clean image $\mathbf{x}_0$, denoted as $\hat{\mathbf{x}}_0$. The training procedure is proposed in Algorithm 4.

Algorithm 4 Proposed training procedure

Input:clean $\mathbf{x}_0$ and noisy inputs $\mathbf{x}_T$, hyperparameter γ, total timesteps T, learning rate η
Output: trained network $\mathbf{U}_\theta$
repeat
  $t \sim \text{Uniform}(\{1, \dots, T\})$ ▷ sample random timestep
  $\alpha_t = (\frac{t}{T})^\gamma$ ▷ compute coefficient
  $\mathbf{x}_t = (1 - \alpha_t)\mathbf{x}_0 + \alpha_t \mathbf{x}_T$ ▷ generate intermediate image
  $\hat{\mathbf{x}}_0 = \mathbf{U}_\theta(\mathbf{x}_t, t)$ ▷ predict clean image
  $\mathcal{L} = \mathcal{L}_{\text{TOTAL}}(\hat{\mathbf{x}}_0, \mathbf{x}_0)$ ▷ compute loss
  $\nabla_\theta \mathcal{L}$ ▷ backpropagate gradients
  $\theta \leftarrow \theta - \eta \cdot \nabla_\theta \mathcal{L}$ ▷ update network parameters
until convergence

First, we set the hyperparameter γ and sample a clean image $\mathbf{x}_0$ and its noisy counterpart

$\mathbf{x}_T$ from the dataset. We then randomly select a timestep t and compute the interpolation weight α_t. Using this coefficient, we generate the intermediate noisy image $\mathbf{x}_t$ as a linear combination of $\mathbf{x}_0$ and $\mathbf{x}_T$. The model $\mathbf{U}_\theta$ takes $\mathbf{x}_t$ and t as inputs and returns the predicted clean image $\hat{\mathbf{x}}_0$. Finally, we compute the loss between $\hat{\mathbf{x}}_0$ and the ground truth $\mathbf{x}_0$, and update the model parameters via gradient descent.

4.2 DIFFUSION PROCESS

As discussed in Chapter 2, standard diffusion models progressively add Gaussian noise to an image until it becomes pure noise. Traditional image-conditioned diffusion models apply the condition only during the reverse process. However, this leaves the forward process unchanged, still relying on Gaussian noise addition.

As discussed in Chapter 1, in image denoising problems, where the noise present in real-world data usually does not follow a normal distribution, regulating noise addition becomes challenging. In addition, the standard approach may introduce irrelevant residual information when applied to image denoising.

To address this limitation, we redefine the image-conditioned diffusion process to better reflect LDCT image degradation, which does not follow a normal distribution. Rather than injecting synthetic noise into the NDCT image, we propose a redefinition of the forward process as a gradual addition of residual noise in multiple steps. This approach starts with the NDCT image and gradually transforms it into its corresponding LDCT version. Figure 14 illustrates this process.

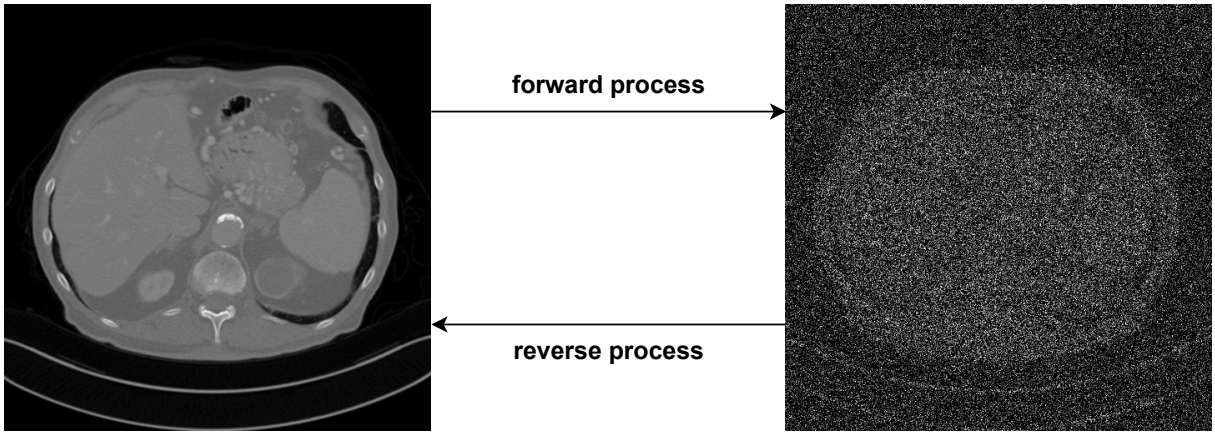


Figure 14 – Illustration of the redefined image-conditioned diffusion process. In the forward process, a clean NDCT image is progressively transformed into its corresponding LDCT version. The reverse process aims to recover the NDCT image from the LDCT input through a learned denoising model.

4.2.1 Proposed power-law noise schedule

The primary difference between our approach and conventional DDPMs, which were presented in Chapter 2, consists in the noise schedule. Instead of using a fixed variance schedule such as those presented in Section 2.4, we employ a power-law based interpolation scheme that ensures a flexible transition between NDCT and LDCT images. Unlike conventional schedules that rely on explicitly predefined noise levels, our method applies a linear combination approach that ensures that, at timestep $t = 0$, the image is the original NDCT image, and at timestep $t = T$, the image corresponds to the LDCT image. Figure 15 illustrates how this process works.

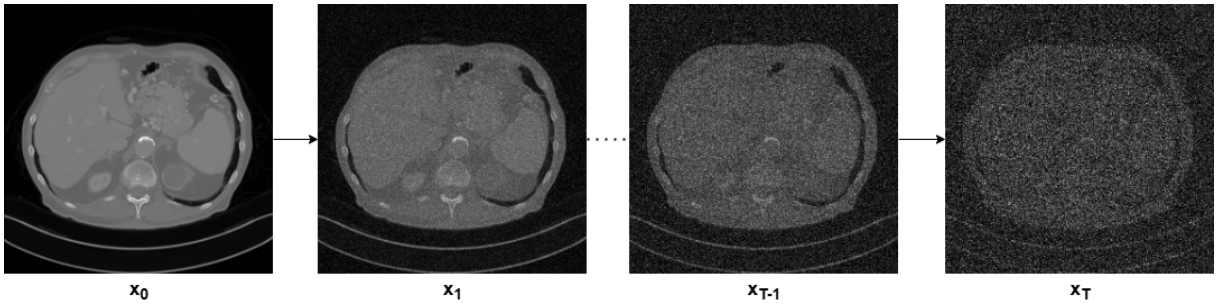


Figure 15 – Example of the redefined noise schedule. The process starts with the NDCT image at timestep $t = 0$ and finishes with the LDCT image at timestep $t = T$. The intermediate images $\mathbf{x}_t$ are interpolated versions of these two inputs generated through a linear combination scheme.

In traditional diffusion models, noisy samples at timestep t are generated as

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\,\boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I}), \tag{4.1}$$

where $\mathbf{x}_0$ is the clean image and $\boldsymbol{\epsilon}$ represents Gaussian noise. This formulation assumes a synthetic, isotropic noise model, which may not accurately reflect the noise characteristics present in real-world LDCT data. Inspired by some prior works that have explored image-conditioned diffusion (HO et al., 2022; CHUNG et al., 2022), we choose to explicitly integrate the real noisy image $\mathbf{x}_T$ into the forward process, replacing Gaussian noise with a known noisy observation. A naive attempt could be formulated as:

$$\mathbf{x}_t = \sqrt{(1-\alpha_t)}\mathbf{x}_0 + \sqrt{\alpha_t}\mathbf{x}_T, \tag{4.2}$$

in analogy to Equation 4.1. However, this introduces residual artifacts because $\sqrt{(1-\alpha_t)} + \sqrt{\alpha_t} \neq 1$, which violates interpolation consistency (GAO et al., 2023). To address this issue, we adopt a simpler interpolation scheme (SONG; MENG; ERMON, 2020)

$$\mathbf{x}_t = (1-\alpha_t)\mathbf{x}_0 + \alpha_t\mathbf{x}_T, \tag{4.3}$$

where $\alpha_t \in (0, 1)$ determines the proportion of noise added to the clean image at timestep $t \in \mathbb{Z}$, $1 < t < T$, with $T \in \mathbb{N}$ being the total number diffusion steps. Given a pair of images $(\mathbf{x}_0, \mathbf{x}_T)$, where $\mathbf{x}_0 \in \mathbb{R}^{H \times W}$ is the clean NDCT image and $\mathbf{x}_T \in \mathbb{R}^{H \times W}$ is the corresponding noisy LDCT image, α_t is defined as:

$$\alpha_t = \left(\frac{t}{T}\right)^{\gamma} \tag{4.4}$$

Small values of α result in images closer to $\mathbf{x}_0$, while larger values produce images closer to $\mathbf{x}_T$. That way, as α increases with t, the images of intermediate timesteps become increasingly closer to the noisy input. Figure 16 shows the α_t progression for different choices of $\gamma \in \mathbb{R}$.

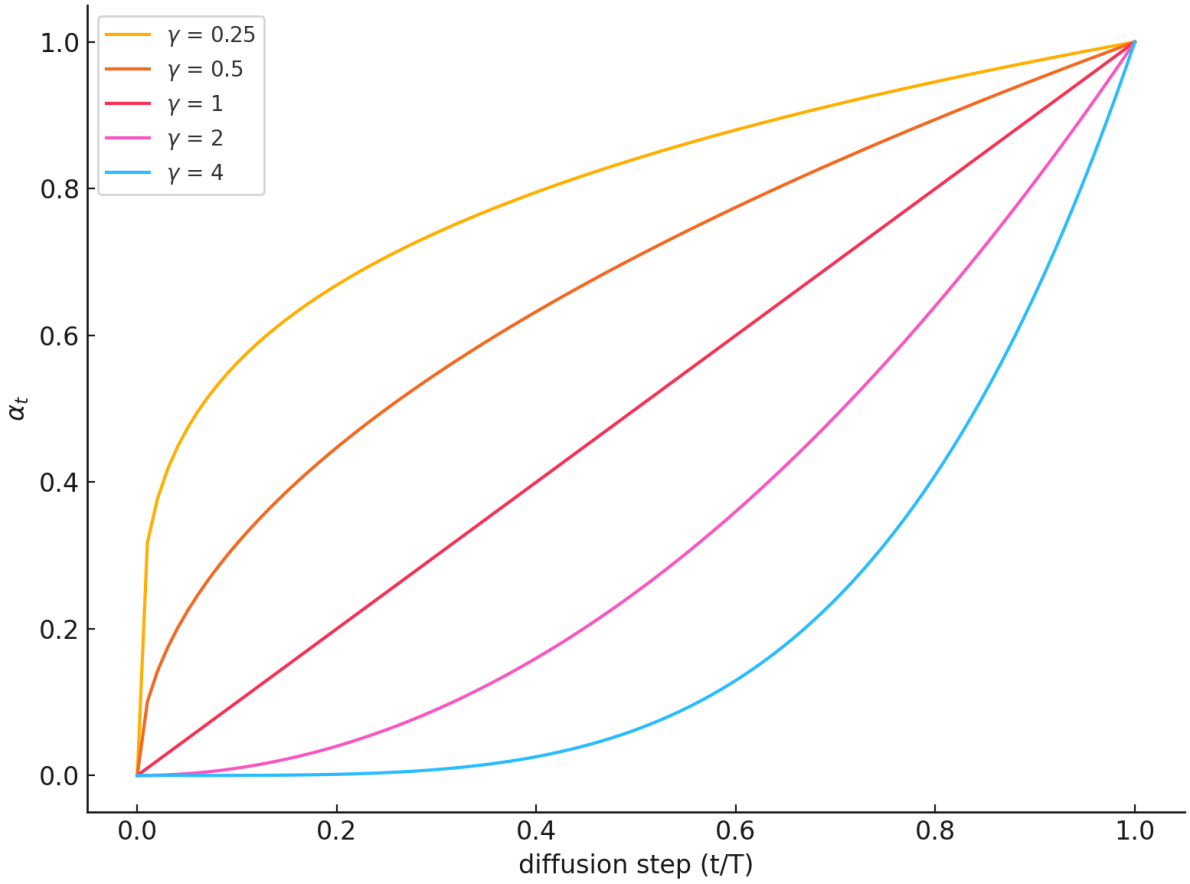


Figure 16 – Behavior of α_t throughout diffusion steps under the proposed γ-parameterized noise schedule. This schedule is the core of our method: it allows dynamic control over the pace at which the noisy input $\mathbf{x}_T$ is progressively added into the clean input $\mathbf{x}_0$. $\gamma < 1$ lead to a faster increase of α_t, which results in noisier intermediate images $\mathbf{x}_t$ at the beginning of the forward process. On the other hand, $\gamma > 1$ result in a delayed growth of α_t, causing a slower transformation in the noisy input.

4.2.2 The proposed γ factor

We introduce a power-law schedule modulated by a scalar hyperparameter γ, which defines the rate of the transition from NDCT to LDCT images over the diffusion process. When $\gamma = 1$, the interpolation follows a linear trajectory. For $\gamma < 1$, degradation is accelerated in early steps, leading to noisier intermediate images, whereas for $\gamma > 1$, the transition starts more softly and intensifies in later steps. This flexibility allows the model to adapt to different noise characteristics observed in real-world low-dose CT scenarios.

Coincidentally, this formulation provides a natural bridge to the theory of Curriculum Learning (BENGIO et al., 2009), where a model is trained starting with simpler examples and is gradually introduced to more complex ones. In our case, setting $\gamma > 1$ produces a curriculum schedule, where early training steps involve nearly clean images, and noise increases progressively. Conversely, setting $\gamma < 1$ induces a Reverse Curriculum Learning regime (WEINSHALL; COHEN; AMIR, 2018), where the model is presented with noisier samples from the beginning, potentially encouraging robustness early on. Figure 17 visually demonstrates the influence of different γ values on the progression of noise over time.

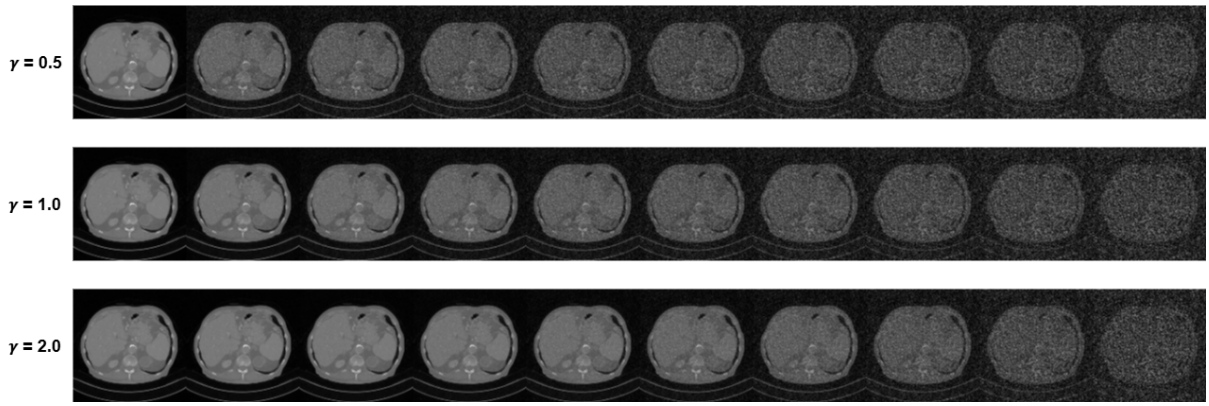


Figure 17 – Example of samples generated with different values of γ. Each row corresponds to a different γ: the top row represents $\gamma = 0.5$, the middle row is the traditional linear schedule $\gamma = 1.0$ and the bottom row is equivalent to $\gamma = 2.0$, respectively. Compared to the linear schedule, the lower value of γ concentrate noise addition at earlier stages, while the higher value produces a more aggressive transition at the end of the process.

4.2.3 Loss optimization

The choice of loss is one of the main aspects to be considered when training diffusion models (KINGMA et al., 2021). Studies conducted by Lin e Yang (2025) and Berrada et al. (2025) show that, in terms of diffusion conditional guidance, the MSE objective works worse

than the perceptual objective. In addition, in the context of CT denoising (ALMEIDA, 2024), pixel-wise losses can cause the MSE effect: a phenomenon that compromises the reconstruction of fine details, leading to over-smoothed images that lack sharpness in textures and edges (WOLTERINK et al., 2017; YU et al., 2017; LI et al., 2020). This effect can also influence the quantitative analysis by resulting in intrinsically higher metric values (YANG et al., 2018).

For that reason, in order to optimize the denoising performance, we adopted a weighted combination of both the MSE loss (WANG; BOVIK, 2009) and the VGG perceptual loss (JOHNSON; ALAHI; FEI-FEI, 2016) for our training. The total loss used is defined as:

$$\mathcal{L}_{\text{TOTAL}} = \lambda_1 \mathcal{L}_{\text{MSE}} + \lambda_2 \mathcal{L}_{\text{VGG}}, \tag{4.5}$$

where $\mathcal{L}_{MSE}$ refers to the mean squared error and $\mathcal{L}_{VGG}$ is the perceptual loss calculated on feature representations extracted from a pre-trained VGG-19 (SIMONYAN; ZISSERMAN, 2014). The hyperparameters λ_1 e λ_2 balance the contribution of the perceptual loss relative to the pixel-wise loss, and are set to 1 and 10^{-3}, respectively.

4.3 INFERENCE

The effectiveness of the denoising process is highly dependent on the sampling strategy adopted by the diffusion model (MA et al., 2025). In conventional DDPMs, the inference stage consists of a sequential denoising procedure that progressively reconstructs a clean image $\mathbf{x}_0$ from pure noise $\mathbf{x}_T \sim \mathcal{N}(0, \mathbf{I})$. This process is inherently stochastic and relies on a series of Gaussian-based reverse transitions, as formulated in Algorithm 2. However, in our proposed LDCT denoising formulation such stochasticity may introduce unnecessary variance.

To address this limitation, we adopt the deterministic sampling proposed in Algorithm 3 of Cold Diffusion (BANSAL et al., 2022). Cold Diffusion redefines the degradation process in diffusion models as an arbitrary, deterministic transformation D, and trains a restoration operator R to reverse it. The key insight from Cold Diffusion is that injecting a controlled amount of degradation even during inference enhances sampling stability and reduces error accumulation. The sampling strategy adopted by our proposed method is described in Algorithm 5 and formalized in Equations 4.6 e 4.7.

The sampling process begin with with a noisy image $\mathbf{x}_T$, sampled from the data distribution. At each timestep t, the algorithm computes a modulation factor $\alpha_t = \left(\frac{t}{T}\right)^{\gamma}$, which controls the relative contribution of the initial noisy image and the predicted clean image $\hat{\mathbf{x}}_0$. Then,

Algorithm 5 Adapted deterministic sampling

Input: noisy input $\mathbf{x}_T$, hyperparameter γ
Output: denoised image $\mathbf{x}_0$
for $t = T, \ldots, 1$ **do**
    $\alpha_t = \left(\frac{t}{T}\right)^{\gamma}$
    $\alpha_{t-1} = \left(\frac{t-1}{T}\right)^{\gamma}$
    $\hat{\mathbf{x}}_0 = \mathbf{U}_\theta(\mathbf{x}_t, t)$
    $\tilde{\mathbf{x}}_t = (1 - \alpha_t)\hat{\mathbf{x}}_0 + \alpha_t \mathbf{x}_T$
    $\tilde{\mathbf{x}}_{t-1} = (1 - \alpha_{t-1})\hat{\mathbf{x}}_0 + \alpha_{t-1}\mathbf{x}_T$
    $\mathbf{x}_{t-1} = \mathbf{x}_t - \tilde{\mathbf{x}}_t + \tilde{\mathbf{x}}_{t-1}$
end for

an intermediate target state $\tilde{\mathbf{x}}_t$ is computed as a linear interpolation between $\hat{\mathbf{x}}_0$ and $\mathbf{x}_T$, modulated by α_t. The denoised estimate at timestep $t-1$, $\mathbf{x}_{t-1}$, is then obtained by adjusting the previous state $\mathbf{x}_t$ according to the change in interpolation from $\tilde{\mathbf{x}}_t$ to $\tilde{\mathbf{x}}_{t-1}$. The process iterates until $t = 1$, at which point $\mathbf{x}_0$ represents the final denoised output.

4.3.1 Deterministic Sampling Strategy

Our sampling procedure builds upon the improved deterministic strategy introduced in the Cold Diffusion framework (BANSAL et al., 2022). In this formulation, the degradation operator D is used to generate intermediate states between the predicted clean image $\hat{\mathbf{x}}_0$ and the degraded observation $\mathbf{x}_T$, guiding the iterative reconstruction process.

In our implementation, we define the degradation operator D as a weighted combination of the noisy observation $\mathbf{x}_T$ and the predicted clean image $\hat{\mathbf{x}}_0$. In conformity with Equation 4.3, the interpolation terms used in Algorithm 5 can be formally expressed as:

$$D(\hat{\mathbf{x}}_0, t) = \mathbf{x}_t = (1 - \alpha_t)\hat{\mathbf{x}}_0 + \alpha_t \mathbf{x}_T, \qquad \alpha_t = \left(\frac{t}{T}\right)^{\gamma} \tag{4.6}$$

$$D(\hat{\mathbf{x}}_0, t-1) = \mathbf{x}_{t-1} = (1 - \alpha_{t-1})\hat{\mathbf{x}}_0 + \alpha_{t-1}\mathbf{x}_T, \qquad \alpha_{t-1} = \left(\frac{t-1}{T}\right)^{\gamma} \tag{4.7}$$

This strategy customizes the restoration step as an interpolation between the current estimate and the noisy input, progressively refining the result while preserving global structure. The deterministic nature of this sampling allows for more stable reconstructions and better alignment with our training objective.

5 EXPERIMENTS DESIGN

This chapter presents the overall experimental pipeline proposed to evaluate our method. The dataset used is introduced in Section 5.1. Section 5.2 details the choices of our network's architectural design, as well as the hyperparameters γ and T. Then, in Section 4.2.3 we discussed loss optimization for denoising of medical images. Section 5.3 presents the experimental environment used. Finally, Section 5.4 describes and justifies the metrics used to evaluate the results obtained.

5.1 MAYO CHALLENGE DATASET

In our experiments, we used the 2016 Low-Dose CT Grand Challenge dataset (MCCOLLOUGH et al., 2017), sponsored by the American Association of Physicists in Medicine (AAPM), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), and Mayo Clinic, for training and inference. The data used in this study were provided by the Mayo's CT Clinical Innovation Center and consist of abdominal CT scans of 30 patients acquired in the portal venous phase with the Siemens SOMATOM Flash scanner (AAPM, 2016). The provided data included cases of both benign and metastatic lesions, as well as cases without lesions.

The full-dose data were acquired at 120 kV and 200 quality reference mAs (QRM), while the simulated low-dose data correspond to 120 kV and 50 QRM, that is, a quarter of the full dose (CHRISTE et al., 2013). To simulate low-dose conditions, Poisson noise was inserted into the sinogram (projection) data, reaching a noise level corresponding to 25% of the original dose. Both full and quarter-dose images were reconstructed using filtered backprojection (FBP) (AAPM, 2016).

The full dataset is divided into training and testing subsets. The training set comprises 10 patients, with both full-dose and quarter-dose images available, while the testing set contains the remaining 20 patients with only quarter-dose images available. The data are organized into four subsets based on slice thickness (1mm and 3mm) and reconstruction kernel (B30 and D45) (AAPM, 2016).

The dataset division ensures a robust evaluation of denoising algorithms under varying noise and resolution conditions. The 1mm slices provide higher spatial resolution but retain more noise, whereas the 3mm slices reduce noise at the expense of lower resolution due to

partial averaging (FORD; DECKER, 2016). Thinner slices are usually preferred for visualizing fine details and structures, while thicker slices can reduce radiation exposure (KATARIA et al., 2020). Additionally, the B30 kernel produces smoother images with suppressed noise but less sharpness, while the sharp D45 kernel enhances structural details at the cost of amplified noise (ZENG et al., 2022). By combining these parameters the dataset allows for a more robust evaluation of denoising algorithms in terms of image quality, noise levels, and spatial resolution.

The Mayo Challenge database continues to be widely used by scientists world-wide and have become a de facto standard set of reference data for use in developing and evaluating CT reconstruction and denoising techniques (Mayo Clinic, 2016). Figure 18 shows some representative examples of images from this dataset.

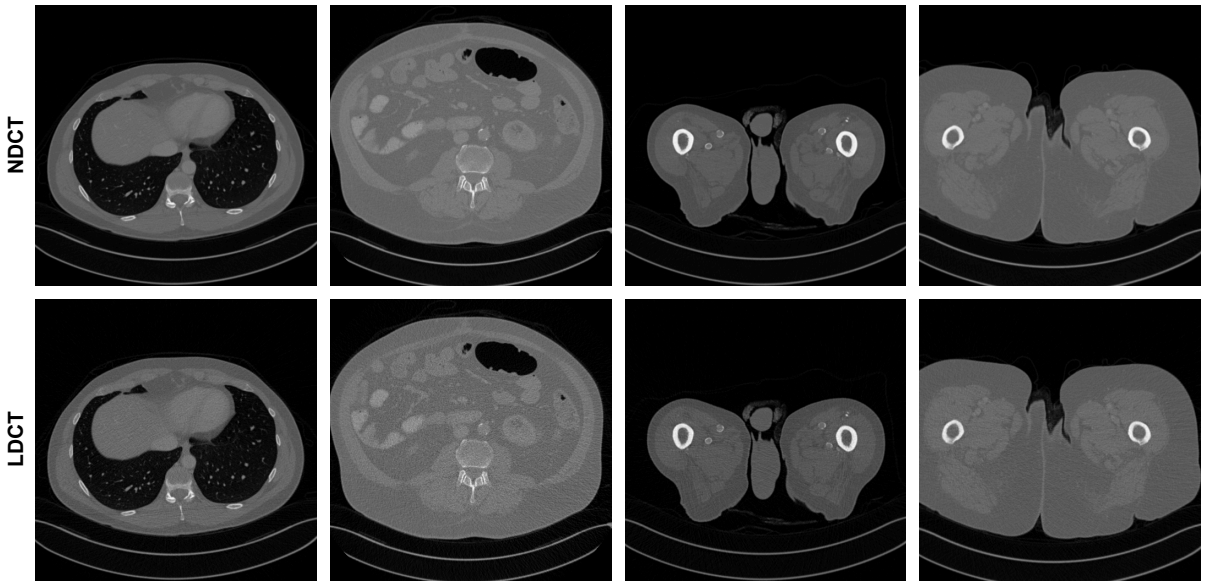


Figure 18 – Example of images from the 2016 Low Dose CT Grand Challenge dataset. The top row represents the CT scans obtained with the use of full radiation dose (NDCT), while the bottom row displays their corresponding low-dose counterparts (LDCT). The LDCT condition was simulated by injecting Poisson noise into the projection data, emulating the effects of a noise level corresponding to 25% of the original full dose result.

For our experiments, we used the complete set of full dose and quarter dose images from the 10 patients of the training set. We further selected 8 patients (L067, L109, L143, L192, L286, L291, L310 and L333) for the training phase, 1 patient (L096) for the validation set and 1 patient (L506) for the test set. All four subsets (1mm B30, 1mm D45, 3mm B30, and 3mm D45) of images were included in our training and inference pipeline. We used patches of size 64×64 to train our diffusion network and the full images with resolutions of 512×512 during inference. No additional processing was required, since the the data was already normalized in the adequate range for the purposes of our experiments.

5.2 EXPERIMENT PIPELINE

Following the standard training pipeline of a diffusion models (HO; JAIN; ABBEEL, 2020), we employed a U-Net (RONNEBERGER; FISCHER; BROX, 2015) architecture as the backbone of our reverse diffusion network. While a variety of arhitectural designs are available, ranging from simple convolutional networks to complex attention-based models, we opted for a simpler and more computationally efficient approach. Our U-Net consists of four downsampling blocks, a central bottleneck block, four upsampling blocks, and a final output convolutional layer.

The experimental setup was designed to evaluate the model's denoising performance under varying hyperparameter conditions. Specifically, we focused on two aspects: the effect of the γ parameter in the noise schedule and the total number of diffusion steps T.

For the γ parameter, which modulates the intensity of the noise addition in the forward process, we tested three configurations:

- $\gamma = 1.0$, corresponding to a linear noise schedule;
- $\gamma = 0.5$, leading to a slower addition in early steps;
- $\gamma = 2.0$, resulting in a more aggressive addition concentrated in later stages.

In addition to γ, we also explored variations on the total number diffusion steps. We tested $T = 10$, $T = 50$, and $T = 100$, using the same value of T during both training and inference. The computational effort increased consistent with the number of diffusion steps: training with $T = 10$ required approximately 68 hours, $T = 50$ took around 81 hours, and $T = 100$ required about 97 hours. This experimental framework allowed us to analyze how both the noise schedule and the diffusion steps affect the denoising quality and computational cost of LDCT reconstruction.

5.3 IMPLEMENTATION DETAILS

All experiments were conducted in a Linux-based environment running on a server equipped with an NVIDIA A100 GPU with 80 GB of memory. The entire diffusion model was implemented in Python v3.9.6 using the PyTorch deep learning framework (v2.2.0) (PASZKE et al., 2019). GPU acceleration was enabled through CUDA v12.1 and cuDNN v8.

We trained all models using the AdamW (Adaptive Moment Estimation with Weight Decay) (KINGMA; BA, 2014) optimizer, whose hyperparameters were set to $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The learning rate was initially set to 3×10^{-4}, and a ReduceLROnPlateau scheduler was employed to adapt its value dynamically based on the validation loss, reducing it when performance stagnated. We performed 3000 training iterations, with a validation step executed every 100 epochs.

To ensure a fair comparison across configurations, all experiments were conducted using the same training, validation, and test split The model's performance was quantitatively evaluated using standard image quality metrics, which are detailed in Section 5.4.

5.4 QUANTITATIVE EVALUATION

Quantitative metrics are important tools to evaluate the performance of neural network models. The following commonly used objective image quality assessment metrics were employed to evaluate the denoising performance of our framework: Structure Similarity Index Measure (SSIM), Peak Signal-to-Noise Ratio (PSNR), and Root Mean Squared Error (RMSE).

SSIM (WANG et al., 2004) measures the degradation of structural information between two images. It aims to quantify perceived visual quality by comparing luminance (l), contrast (c), and structural similarity (s) between the predicted image $\mathbf{y}$ and the reference image $\mathbf{x}$. The SSIM index is defined as:

$$\text{SSIM}(\mathbf{x}, \mathbf{y}) = [l(\mathbf{x}, \mathbf{y})]^{\alpha} \cdot [c(\mathbf{x}, \mathbf{y})]^{\beta} \cdot [s(\mathbf{x}, \mathbf{y})]^{\gamma}, \tag{5.1}$$

where α, β, and γ are parameters to adjust the relative importance of each component (typically set to 1). The individual components are defined as

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

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

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

where $C_1 = (K_1 L)^2$, $C_2 = (K_2 L)^2$, and $C_3 = C_2/2$ are constants to stabilize the division when denominators are small. L is the dynamic range of the image pixel values, usually $L = 255$ or $L = 1$ for normalized images, and typically $K_1 = 0.01$ and $K_2 = 0.03$. The final SSIM

formula, when $\alpha = \beta = \gamma = 1$, becomes:

$$\text{SSIM}(\mathbf{x}, \mathbf{y}) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \tag{5.2}$$

where μ_x, μ_y are the local means of $\mathbf{x}$ and $\mathbf{y}$, σ_x, σ_y are the local standard deviations, and σ_{xy} is the local cross-covariance between $\mathbf{x}$ and $\mathbf{y}$. SSIM values range from -1 to 1, with 1 indicating perfect similarity.

PSNR measures the fidelity of a signal with respect to noise. Higher PSNR values indicate better quality reconstructions. It is derived from the Mean Squared Error (MSE) and is defined as:

$$\text{PSNR}(\mathbf{x}, \mathbf{y}) = 10 \cdot \log_{10} \left(\frac{MAX_I^2}{\text{MSE}(\mathbf{x}, \mathbf{y})} \right), \tag{5.3}$$

where MAX_I is the maximum possible pixel value of the image.

RMSE is a widely used metric that quantifies the difference between two signals. It computes the square root of the average of the squared differences between corresponding pixel values:

$$\text{RMSE}(\mathbf{x}, \mathbf{y}) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - y_i)^2}, \tag{5.4}$$

where N is the total number of pixels. Unlike SSIM and PSNR, RMSE is minimized—the lower the value, the better the reconstruction quality.

6 RESULTS

In this chapter, we present and analyze the results obtained with the proposed diffusion-based denoising method including the ablation study of varying the hyperparameter γ within the proposed noise scheduler associated with the DDPM algorithm. To verify the reconstruction quality of the low-dose CT images, our evaluation was conducted using the standard image quality metrics SSIM, PSNR, and RMSE. We adopted the U-Net presented in Section 5.2 as our reference baseline. The goal of these experiments is to validate the effectiveness of our proposed conditioning denoising process.

In the following sections, we will use the term FBP referring to the input image reconstructed using the classic Filtered BackProjection algorithm, which is the low-dose CT image itself. The U-Net corresponds to the baseline denoising network, trained to directly map noisy inputs to clean outputs. The notations DDPM-10, DDPM-50, and DDPM-100 represent the proposed diffusion-based denoising model using $T = 10$, $T = 50$, and $T = 100$ steps, respectively.

6.1 IMPACT OF THE POWER-LAW NOISE SCHEDULER PARAMETER γ

This section presents a comparative evaluation of the denoising performance for different values of the parameter γ, specifically $\gamma = 0.5$, $\gamma = 1.0$, and $\gamma = 2.0$. For each subset of the dataset, we report the SSIM, PSNR, and RMSE metrics obtained by different methods, including the FBP and U-Net baselines, as well as DDPM-based models with varying numbers of diffusion steps. These results allow us to analyze how different choices of γ influence the denoising quality across configurations.

6.1.1 Subset 1mm B30

This section presents the results obtained with all evaluated methods trained on the 1mm B30 subset of the 2016 Low-Dose CT Grand Challenge dataset. Tables 1, 2, and 3 show the metric results obtained for this subset, considering the three different values of $\gamma = 0.5$, 1.0, and 2.0.

6.1.1.1 *Denoising performance using* $\gamma = 0.5$

Table 1 shows the performance metrics for the experiments using $\gamma = 0.5$. The diffusion-based models using the proposed power-law noise scheduler outperformed the baseline in most configurations. Notably, the DDPM-10 model achieved the best overall results, with SSIM $= 0.8582$, PSNR $= 28.2766$, and RMSE $= 15.8037$. In contrast, the DDPM-100 model, which uses a larger number of diffusion steps, yielded inferior values even when compared to the U-Net baseline.

Table 1 – Average values of PSNR, SSIM, and RMSE computed on the test set for each configuration evaluated with $\gamma = 0.5$.

Method	SSIM ↑	PSNR ↑	RMSE ↓
FBP	0.8246	24.4688	24.6370
U-Net	0.8404	26.2271	19.9204
DDPM-10	**0.8582**	**28.2766**	**15.8037**
DDPM-50	0.8556	27.8958	16.5160
DDPM-100	0.8401	25.7159	22.2768

6.1.1.2 *Denoising performance using* $\gamma = 1.0$

Table 2 shows the results for $\gamma = 1.0$. Here, the diffusion-based models using the proposed power-law noise scheduler outperformed the baseline in all configurations. Among them, the DDPM-50 model obtained the best performance in terms of SSIM (0.8575), PSNR (28.0442), and RMSE (16.2225). The DDPM-10 model showed a good value of SSIM (0.8566), although with slightly lower performance if compared to $\gamma = 0.5$. On the other hand, the DDPM-100 model improved substantially.

Table 2 – Average values of PSNR, SSIM, and RMSE computed on the test set for each configuration evaluated with $\gamma = 1.0$.

Method	SSIM ↑	PSNR ↑	RMSE ↓
FBP	0.8246	24.4688	24.6370
U-Net	0.8404	26.2271	19.9204
DDPM-10	0.8566	27.6151	17.2269
DDPM-50	**0.8575**	**28.0442**	**16.2225**
DDPM-100	0.8535	27.8176	16.6427

6.1.1.3 *Denoising performance using* $\gamma = 2.0$

For $\gamma = 2.0$, as shown in Table 3, the performance of the diffusion-based models using the proposed power-law noise scheduler significantly outperformed the reference methods FBP and U-Net in all metrics. Despite a slight performance drop, DDPM-50 maintained consistent results. The DDPM-10 model presented the best values for all metrics in all tested configurations. Notably, the results of the DDPM-100 model also benefited from increased γ, achieving its best results so far, with SSIM $= 0.8574$, PSNR $= 28.0910$, and RMSE $= 16.1563$.

Table 3 – Average values of PSNR, SSIM, and RMSE computed on the test set for each configuration evaluated with $\gamma = 2.0$.

Method	SSIM ↑	PSNR ↑	RMSE ↓
FBP	0.8246	24.4688	24.6370
U-Net	0.8404	26.2271	19.9204
DDPM-10	**0.8600**	**28.4501**	**15.4728**
DDPM-50	0.8565	27.7689	16.7264
DDPM-100	0.8574	28.0910	16.1563

6.1.2 Subset 1mm D45

In this section we present the quantitative results of all evaluated methods trained in the subset 1mm D45 of the 2016 Low-Dose CT Grand Challenge dataset. Tables 4, 5, and 6 show the quantitative results for this subset, considering the three different values of $\gamma = 0.5$, 1.0, and 2.0, respectively. Beforehand, the DDPM-based methods consistently outperform both the FBP and the U-Net baseline in all configurations.

6.1.2.1 *Denoising performance using* $\gamma = 0.5$

For $\gamma = 0.5$, as shown in Table 4, the DDPM-10 algorithm using the proposed power-law noise scheduler achieves the best overall performance, with an SSIM of 0.7974, PSNR of 22.0386, and RMSE of 32.3823.

Table 4 – Average values of PSNR, SSIM, and RMSE computed on the test set for each configuration evaluated with $\gamma = 0.5$.

Method	SSIM ↑	PSNR ↑	RMSE ↓
FBP	0.7582	17.4163	55.0131
U-Net	0.7810	19.7124	42.1110
DDPM-10	**0.7974**	**22.0386**	**32.3823**
DDPM-50	0.7958	21.6794	33.6719
DDPM-100	0.7932	21.7063	33.6637

6.1.2.2 Denoising performance using $\gamma = 1.0$

With $\gamma = 1.0$, DDPM-10 remains the most effective configuration, improving its SSIM and PSNR to 0.7979 and 22.0713, respectively, and reducing the RMSE to 32.2554. On the other hand, DDPM-50 and DDPM-100 do not benefit from the increased γ, and their results worsen when compared to the previous configuration.

Table 5 – Average values of PSNR, SSIM, and RMSE computed on the test set for each configuration evaluated with $\gamma = 1.0$.

Method	SSIM ↑	PSNR ↑	RMSE ↓
FBP	0.7582	17.4163	55.0131
U-Net	0.7810	19.7124	42.1110
DDPM-10	**0.7979**	**22.0713**	**32.2554**
DDPM-50	0.7802	21.4759	34.4874
DDPM-100	0.7961	21.6237	34.0327

6.1.2.3 Denoising performance using $\gamma = 2.0$

Table 6 – Average values of PSNR, SSIM, and RMSE computed on the test set for each configuration evaluated with $\gamma = 2.0$.

Method	SSIM ↑	PSNR ↑	RMSE ↓
FBP	0.7582	17.4163	55.0131
U-Net	0.7810	19.7124	42.1110
DDPM-10	**0.8006**	**22.1844**	**31.8486**
DDPM-50	0.7892	21.4153	34.9011
DDPM-100	0.7886	20.8443	36.7960

Table 6 shows the results for $\gamma = 1.0$. The DDPM-10 algorithm using the proposed power-law noise scheduler once again shows the highest performance across all metrics, reaching an

SSIM of 0.8006, PSNR of 22.1844, and RMSE of 31.8486, which are the best overall values in this subset.

6.1.3 Subset 3mm B30

This section presents the quantitative results for the 3mm B30 subset of the 2016 Low-Dose CT Grand Challenge dataset, evaluating the performance of all methods across different γ values.

6.1.3.1 Denoising performance using $\gamma = 0.5$

For $\gamma = 0.5$, whose results are shown in Table 7, the diffusion-based models using the proposed power-law noise scheduler consistently outperformed both FBP and U-Net baselines. Notably, DDPM-100 achieved the highest SSIM (0.9024) and PSNR (32.0904) while also yielding the lowest RMSE (10.1977), suggesting that a longer diffusion process was beneficial for this subset. DDPM-50 performed well but was still surpassed by DDPM-100 in all metrics.

Table 7 – Average values of PSNR, SSIM, and RMSE computed on the test set for each configuration evaluated with $\gamma = 0.5$.

Method	SSIM ↑	PSNR ↑	RMSE ↓
FBP	0.8759	29.2489	14.2416
U-Net	0.8815	29.9050	13.4720
DDPM-10	0.8899	30.7692	11.8186
DDPM-50	0.8964	30.5294	12.8385
DDPM-100	**0.9024**	**32.0904**	**10.1977**

6.1.3.2 Denoising performance using $\gamma = 1.0$

When increasing γ to 1.0, as presented in Table 8, the DDPM-10 model showed significant improvement, achieving the best SSIM (0.9056) and PSNR (32.1137) among all configurations. DDPM-100 shows a slight drop in performance compared to $\gamma = 0.5$, but still outperforms the baseline model.

Table 8 – Average values of PSNR, SSIM, and RMSE computed on the test set for each configuration evaluated with $\gamma = 1.0$.

Method	SSIM ↑	PSNR ↑	RMSE ↓
FBP	0.8759	29.2489	14.2416
U-Net	0.8815	29.9050	13.4720
DDPM-10	**0.9056**	**32.1137**	**10.2571**
DDPM-50	0.9028	31.6889	10.9820
DDPM-100	0.9044	31.9737	10.3872

6.1.3.3 Denoising performance using $\gamma = 2.0$

Table 9 presents the results obtained at $\gamma = 2.0$. The DDPM-100 algorithm using the proposed power-law noise scheduler regained its advantage, achieving the highest SSIM (0.9102) and PSNR (32.5248), along with the lowest RMSE (9.6830). Meanwhile, the DDPM-10 and DDPM-50 models also maintain stronger performances than the baseline model.

Table 9 – Average values of PSNR, SSIM, and RMSE computed on the test set for each configuration evaluated with $\gamma = 2.0$.

Method	SSIM ↑	PSNR ↑	RMSE ↓
FBP	0.8759	29.2489	14.2416
U-Net	0.8815	29.9050	13.4720
DDPM-10	0.9059	32.5002	9.7152
DDPM-50	0.9011	31.9750	10.2806
DDPM-100	**0.9102**	**32.5248**	**9.6830**

6.1.4 Subset 3mm D45

The following sections covers the 3mm D45 subset of the 2016 Low-Dose CT Grand Challenge dataset, where diffusion models exhibit varying performance depending on γ.

6.1.4.1 Denoising performance using $\gamma = 0.5$

For $\gamma = 0.5$, presented in Table 10, all diffusion-based models outperformed the U-Net baseline in at least one metric. The DDPM-10 algorithm using the proposed power-law noise scheduler had better results for PSNR (25.8926) and RMSE (20.8012). In contrast, DDPM-50 achieved higher SSIM value, but porformed worse in terms of PSNR and RMSE. DDPM-100

achieved the best value of SSIM (0.8403).

Table 10 – Average values of PSNR, SSIM, and RMSE computed on the test set for each configuration evaluated with $\gamma = 0.5$.

Method	SSIM ↑	PSNR ↑	RMSE ↓
FBP	0.8017	21.6048	34.1898
U-Net	0.8375	25.6238	21.3804
DDPM-10	0.8362	**25.8926**	**20.8012**
DDPM-50	0.8387	25.0504	23.3512
DDPM-100	**0.8403**	25.6542	21.4031

6.1.4.2 Denoising performance using $\gamma = 1.0$

The results obtained With $\gamma = 1.0$ are presented in Table 11. DDPM-10 achieved the best SSIM (0.8405), PSNR (25.8318), and RMSE (21.0408). Both DDPM-50 and DDPM-100 models struggled significantly, performing worse than the U-Net in all evaluated metrics.

Table 11 – Average values of PSNR, SSIM, and RMSE computed on the test set for each configuration evaluated with $\gamma = 1.0$.

Method	SSIM ↑	PSNR ↑	RMSE ↓
FBP	0.8017	21.6048	34.1898
U-Net	0.8375	25.6238	21.3804
DDPM-10	**0.8405**	**25.8318**	**21.0408**
DDPM-50	0.8353	25.0886	22.7049
DDPM-100	0.8304	25.0537	23.3485

6.1.4.3 Denoising performance using

At $\gamma = 2.0$, whose results are shown in Table 12, the DDPM-10 algorithm using the proposed power-law noise scheduler again led in SSIM (0.8430), PSNR (25.9625), and RMSE (20.7183). This time, DDPM-50 outperformed the U-Net in terms of SSIM (0.8385) and PSNR (25.6485) when compared to the prior γ results. Even though DDPM-100 performed worse than the baseline, it was still better than the previous γ.

Table 12 – Average values of PSNR, SSIM, and RMSE computed on the test set for each configuration evaluated with $\gamma = 2.0$.

Method	SSIM ↑	PSNR ↑	RMSE ↓
FBP	0.8017	21.6048	34.1898
U-Net	0.8375	25.6238	21.3804
DDPM-10	**0.8430**	**25.9625**	**20.7183**
DDPM-50	0.8385	25.6485	21.4689
DDPM-100	0.8345	25.3909	22.1356

6.2 IMPACT OF THE NUMBER OF DIFFUSION STEPS T

This section presents a comparative evaluation of DDPM models trained with different numbers of diffusion steps, specifically $T = 10$, $T = 50$, and $T = 100$. In addition to reporting the image quality metrics SSIM, PSNR, and RMSE, we analyze the Variational Lower Bound (VLB) curves across diffusion steps for each configuration. In these graphs, higher values indicate higher contributions to the Negative Log-Likelihood (NLL), suggesting that the model struggles to effectively denoise those steps. On the other hand, lower loss curves correspond to tighter VLBs and better approximation of the data likelihood. These analyses help us understand the impact of the choice of T in conjunction with the parameter γ on the denoising quality of the trained models.

6.2.1 Subset 1mm B30

Tables 13, 14 and 15 report the SSIM, PSNR, and RMSE values for different γ values across the DDPM-10, DDPM-50, and DDPM-100 configurations, respectively, trained on the 1mm B30 subset. Figures 19, 20 and 21 display the loss contribution per diffusion step for each of these models.

6.2.1.1 Denoising performance using $T = 10$

Table 13 presents the results for DDPM-10, where the DDPM-10 model presented its best performance with $\gamma = 2.0$, slightly outperforming the other two configurations. The loss curves in Figure 19 show that $\gamma = 1.0$ leads to high losses during the initial steps of the diffusion, which are stabilized later in the process. In contrast, both $\gamma = 0.5$ and $\gamma = 2.0$ yield more

Table 13 – Average SSIM, PSNR, and RMSE metrics on the test set for each γ value tested with DDPM-10.

γ	SSIM ↑	PSNR ↑	RMSE ↓
0.5	0.8582	28.2766	15.8037
1.0	0.8566	27.6151	17.2269
2.0	**0.8600**	**28.4501**	**15.4728**

stable behavior throughout, with the latter performing better in the early steps.

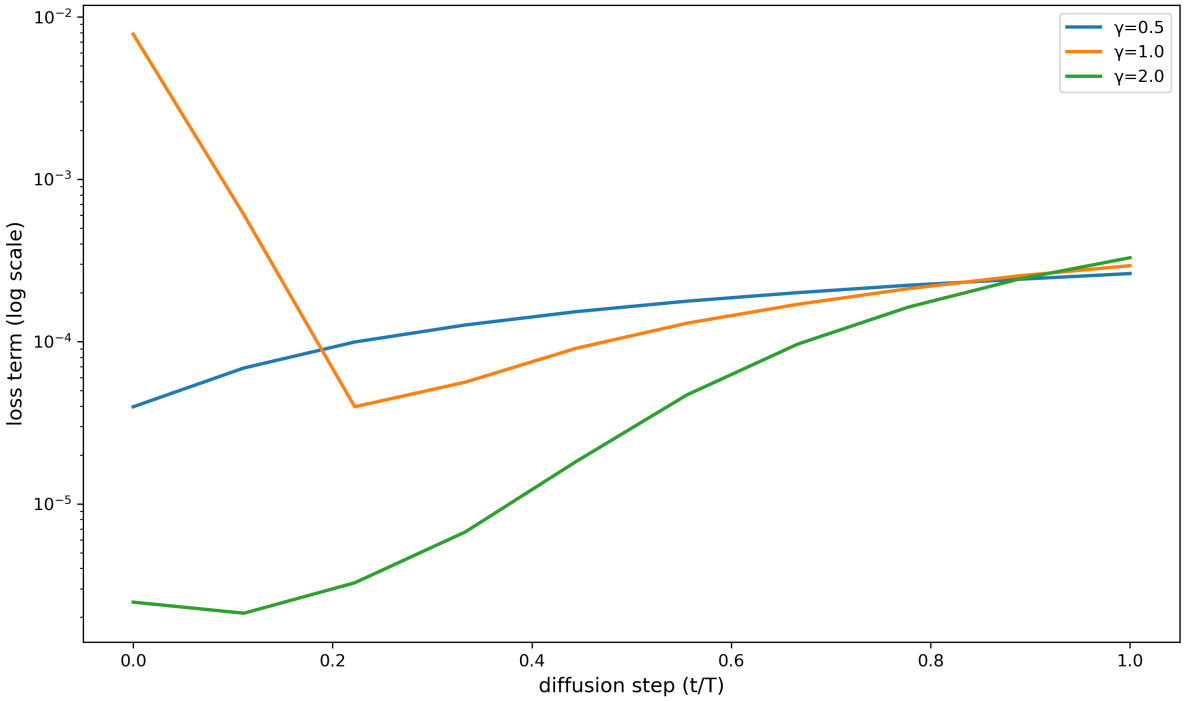


Figure 19 – Loss contribution across diffusion steps for DDPM-10 with different values of γ..

6.2.1.2 Denoising performance using $T = 50$

Table 14 presents the results for DDPM-50, which had its best performance with $\gamma = 1.0$, with minimal variations between SSIM, PSNR, and RMSE compared to the other γ configurations.

Table 14 – Average SSIM, PSNR, and RMSE metrics on the test set for each γ value tested with DDPM-50.

γ	SSIM ↑	PSNR ↑	RMSE ↓
0.5	0.8556	27.8958	16.5160
1.0	**0.8575**	**28.0442**	**16.2225**
2.0	0.8565	27.7689	16.7264

As shown in Figure 20, the curve for $\gamma = 2.0$ exhibits smaller losses in the early diffusion

steps, but the performance worsens in later ones. The model trained with $\gamma = 1.0$, on the other hand, maintained a more stable curve throughout the steps.

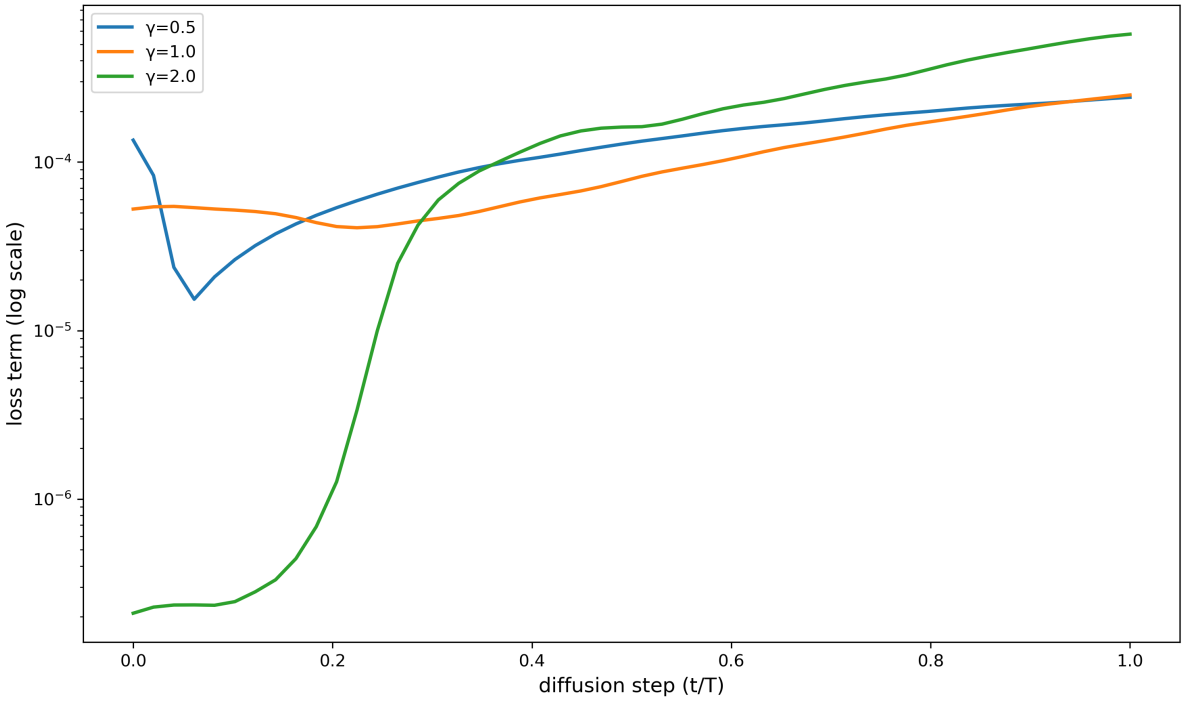


Figure 20 – Loss contribution across diffusion steps for DDPM-50 with different values of γ..

6.2.1.3 Denoising performance using $T = 100$

Table 18 shows the results for DDPM-100. Among all tested configurations in the 1mm B30 subset, DDPM-100 presented the most unstable behavior, with a clear performance improvement as γ increased. As shown in Table 15, the results for $\gamma = 0.5$ were considerably worse than those for $\gamma = 1.0$ and $\gamma = 2.0$.

Table 15 – Average SSIM, PSNR, and RMSE metrics on the test set for each γ value tested with DDPM-100.

γ	**SSIM ↑**	**PSNR ↑**	**RMSE ↓**
0.5	0.8401	25.7159	22.2768
1.0	0.8535	27.8176	16.6427
2.0	**0.8574**	**28.0910**	**16.1563**

The loss curves in Figure 21 show that, for $\gamma = 0.5$, the initial steps dominate the loss, which means the model struggles to learn a balanced denoising process in those timesteps. The $\gamma = 1.0$ configuration shows similar behavior, but with lower loss values. As γ increases

to 2.0, the loss distribution becomes more uniform in the early steps, with the highest losses only appearing later in the process.

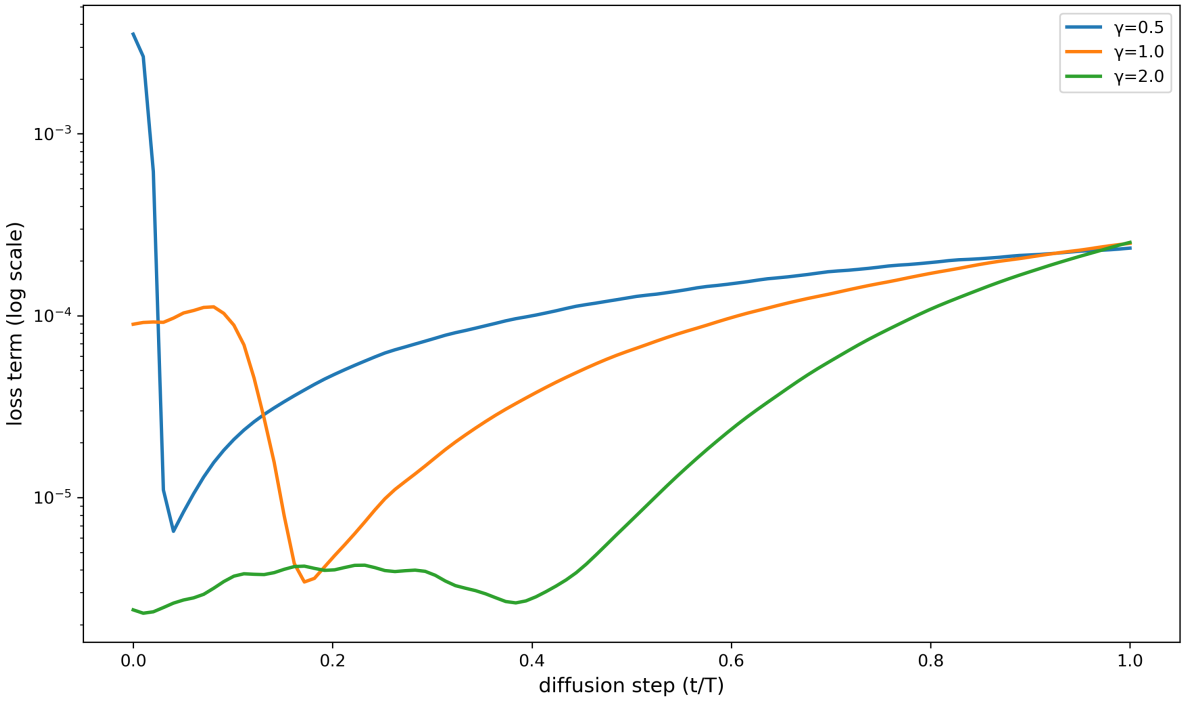


Figure 21 – Loss contribution across diffusion steps for DDPM-100 with different values of γ..

6.2.2 Subset 1mm D45

Tables 16, 17, and 18 report the SSIM, PSNR, and RMSE values for different γ values across the DDPM-10, DDPM-50, and DDPM-100 configurations, respectively, for the 1mm D45 subset. Figures 22, 23 and 24 display the loss contribution per diffusion step for each model.

6.2.2.1 Denoising performance using $T = 10$

Table 16 and Figure 22 present the quantitative results and the evolution of the loss terms, respectively, for the 1mm D45 subset using the DDPM-10 model with different values of γ..

From Table 16, we observe that increasing γ leads to consistent improvements across all metrics. The SSIM increases from 0.7974 to 0.8006, PSNR from 22.0386 to 22.1844, and RMSE decreases from 32.3823 to 31.8486 as γ goes from 0.5 to 2.0.

Figure 22 illustrates how the loss contributions evolve throughout the diffusion steps for

Table 16 – Average SSIM, PSNR, and RMSE metrics on the test set for each γ value tested with DDPM-10.

γ	SSIM ↑	PSNR ↑	RMSE ↓
0.5	0.7974	22.0386	32.3823
1.0	0.7979	22.0713	32.2554
2.0	**0.8006**	**22.1844**	**31.8486**

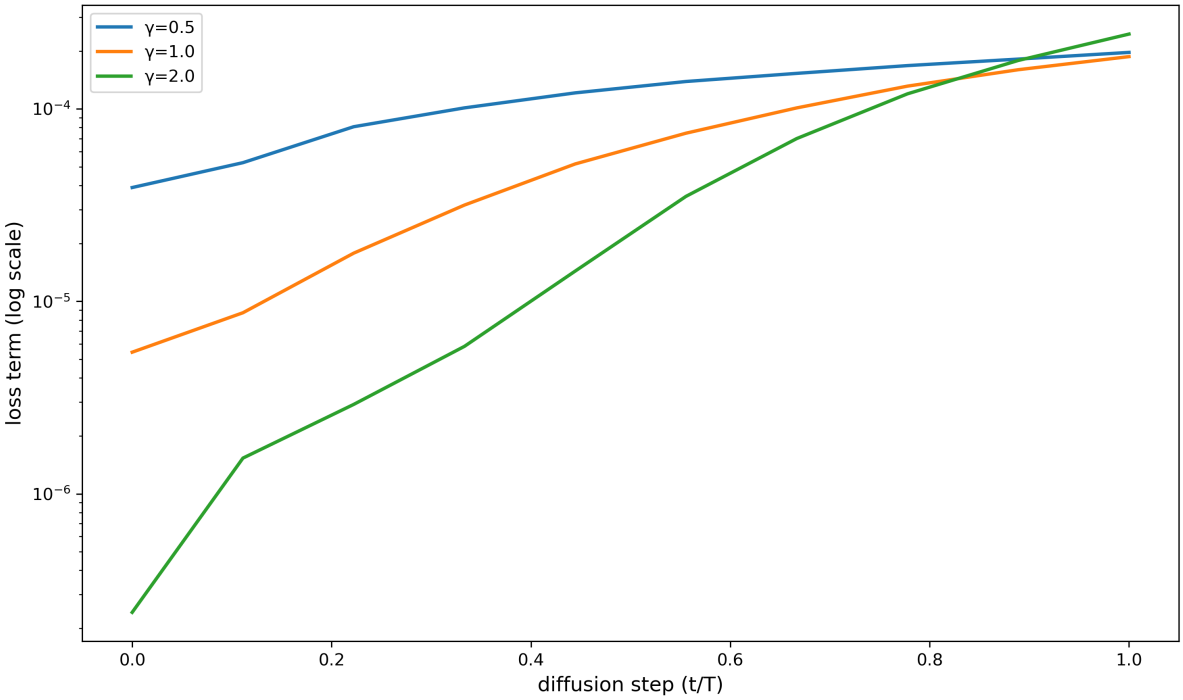


Figure 22 – Loss contribution across diffusion steps for DDPM-10 with different values of γ.

each value of γ. The curves of all configurations behave similarly, with early steps being easier to denoise and higher loss values getting concentrated at the end of the process.

6.2.2.2 Denoising performance using $T = 50$

Table 17 and Figure 23 present the quantitative results and the loss curves obtained using DDPM-50 with different γ values on the 1mm D45 subset.

Table 17 – Average SSIM, PSNR, and RMSE metrics on the test set for each γ value tested with DDPM-50.

γ	SSIM ↑	PSNR ↑	RMSE ↓
0.5	**0.7958**	**21.6794**	**33.6719**
1.0	0.7802	21.4759	34.4874
2.0	0.7892	21.4153	34.9011

Table 17 shows that the results for SSIM, PSNR and RMSE remain relatively stable across

the tested values of γ, with only slight variations. The best SSIM (0.7958), PSNR (21.6794), and RMSE (33.6719) are obtained with $\gamma = 0.5$, while $\gamma = 1.0$ and $\gamma = 2.0$ produce closer results.

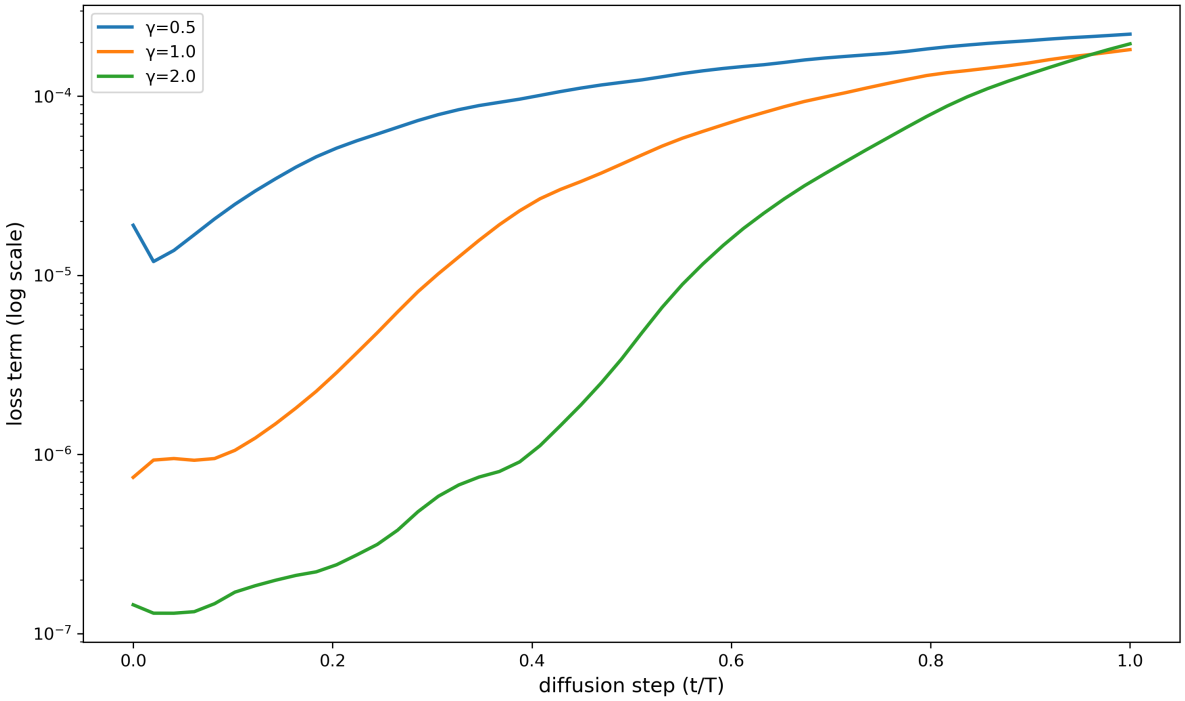


Figure 23 – Loss contribution across diffusion steps for DDPM-50 with different values of γ.

The loss curves illustrated in Figure 23 show a similar behavior to the curves of the previous model, where the loss contribution is more concentrated at the last timesteps.

6.2.2.3 Denoising performance using $T = 100$

Table 18 and Figure 24 present the quantitative results and the VLB loss curves obtained using DDPM-100 with different γ values on the 1mm D45 subset.

Table 18 – Average SSIM, PSNR, and RMSE metrics on the test set for each γ value tested with DDPM-100.

γ	**SSIM ↑**	**PSNR ↑**	**RMSE ↓**
0.5	0.7932	21.7063	33.6637
1.0	0.7961	21.6237	34.0327
2.0	**0.7979**	**21.7770**	**33.3761**

The results of Table 18 show that $\gamma = 2.0$ yielded the best results of SSIM (0.7979) and PSNR (21.7770), while $\gamma = 0.5$ presented the best value of RMSE (33.6637). Even though

improvements were observed as γ increased, here no value of γ stands out as significantly superior.

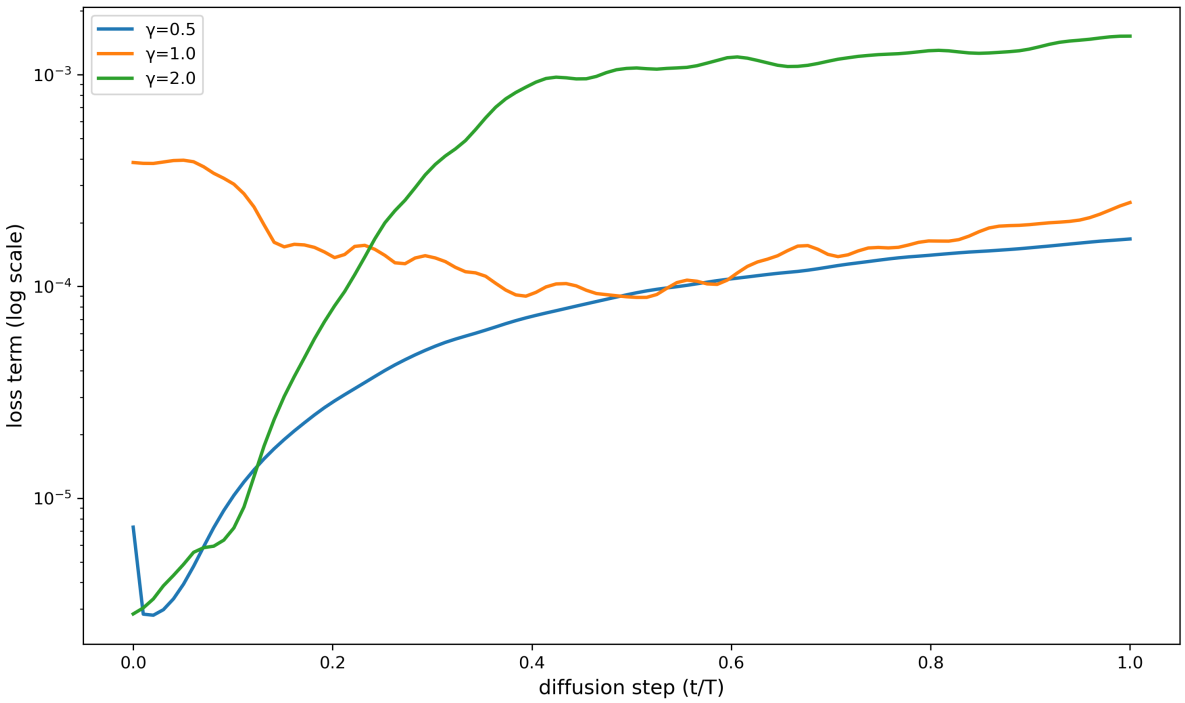


Figure 24 – Loss contribution across diffusion steps for DDPM-100 with different values of γ.

The analysis of Figure 24 shows that for $\gamma = 0.5$ the curve increases steadily throughout the process, indicating that the loss grows as the noise level increases. In contrast, $\gamma = 2.0$ shows a rapid increase in loss during the early steps, which then saturates and remains high, suggesting the model faces greater difficulty at intermediate to late steps due to stronger noise injection. The $\gamma = 1.0$ curve exhibits the most stable behavior, with relatively small variations across steps. Interestingly, it starts slightly higher than the other two, but remains more stable throughout the process.

6.2.3 Subset 3mm B30

Tables 19, 20 and 21 report the SSIM, PSNR, and RMSE values for different γ values across the DDPM-10, DDPM-50, and DDPM-100 configurations, respectively, trained on the 3mm B30 subset. Figures 25, 26 and 27 display the loss contribution per diffusion step for each of these models.

6.2.3.1 *Denoising performance using $T = 10$*

Table 19 presents the quantitative results for DDPM-10 across different γ values. Performance improves consistently as γ increases, with $\gamma = 2.0$ achieving the highest SSIM (0.9059), PSNR (32.5002), and lowest RMSE (9.7152). This suggests that higher γ values enhance denoising performance for shorter diffusion steps in this subset.

Table 19 – Average SSIM, PSNR, and RMSE metrics on the test set for each γ value tested with DDPM-10.

γ	**SSIM ↑**	**PSNR ↑**	**RMSE ↓**
0.5	0.8899	30.7692	11.8186
1.0	0.9056	32.1137	10.2571
2.0	**0.9059**	**32.5002**	**9.7152**

Figure 25 illustrates the loss contribution across timesteps for the DDPM-10 model with varying values of γ. The results indicate that $\gamma = 0.5$ lead to higher loss terms across all steps, with a moderately increasing trend. In contrast, $\gamma = 2.0$ starts with low loss in early steps but exhibits an accentuated increase. The curve for $\gamma = 1.0$ maintains a smoother progression, suggesting a more balanced concentration of the learning signal across the diffusion steps.

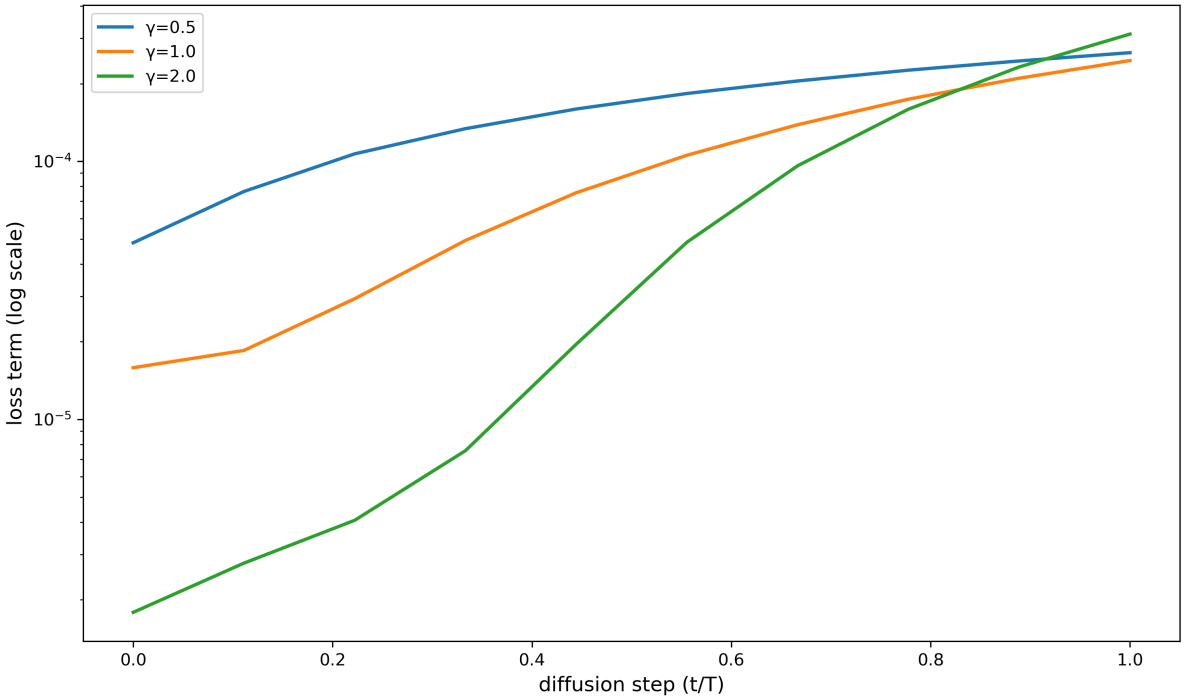


Figure 25 – Loss contribution across diffusion steps for DDPM-10 with different values of γ.

6.2.3.2 *Denoising performance using $T = 50$*

Table 20 shows the results for all γ configurations for DDPM-50. $\gamma = 2.0$ yields the best metrics, with SSIM $= 0.9044$, PSNR $= 31.9737$, and RMSE $= 10.3872$. Once again, higher values of γ tend to perform better when trained with larger values of T.

Table 20 – Average SSIM, PSNR, and RMSE metrics on the test set for each γ value tested with DDPM-50.

γ	SSIM ↑	PSNR ↑	RMSE ↓
0.5	0.8964	30.5294	12.8385
1.0	0.9028	31.6889	10.9820
2.0	**0.9044**	**31.9737**	**10.3872**

Figure 26 illustrates the variational loss terms for the DDPM-50 model under different values of γ. As observed, $\gamma = 0.5$ maintains the highest loss values across almost all diffusion steps, following a nearly monotonic trend. On the other hand, the curve for $\gamma = 2.0$ is significantly less smooth, with extremely lower losses at early steps and rapidly increasing values toward the end. For $\gamma = 1.0$, the loss curve has a slight fluctuation at middle-steps but remains relatively balanced throughout the process.

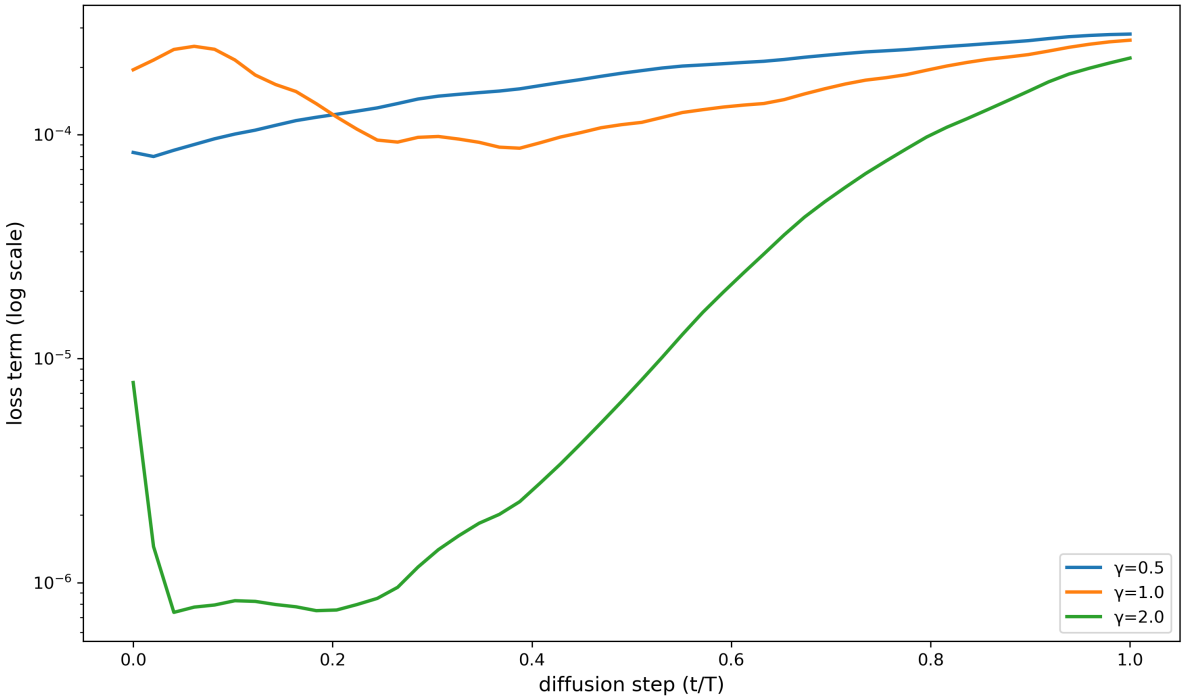


Figure 26 – Loss contribution across diffusion steps for DDPM-50 with different values of γ.

6.2.3.3 *Denoising performance using $T = 100$*

Table 21 presents the results for DDPM-100, which had its best performance at $\gamma = 2.0$, with SSIM=0.9102, PSNR=32.5248, and RMSE=9.6830. The other two configurations, $\gamma = 0.5$ and $\gamma = 1.0$, achieved slightly lower results, with $\gamma = 0.5$ showing the worst values. Although the differences are relatively small, the consistent improvement across all γ variations suggests that higher values of γ can positively impact of models trained in the 3mm B30 subset.

Table 21 – Average SSIM, PSNR, and RMSE metrics on the test set for each γ value tested with DDPM-100.

γ	**SSIM ↑**	**PSNR ↑**	**RMSE ↓**
0.5	0.9024	32.0904	10.1977
1.0	0.9044	31.9737	10.3872
2.0	**0.9102**	**32.5248**	**9.6830**

Figure 27 shows the loss contribution per diffusion step for DDPM-100. The analysis of the curves demonstrate that for $\gamma = 2.0$ the loss values remain low and rapidly increase at later steps. In contrast, for $\gamma = 0.5$, the loss is more uniformly distributed across the entire process. The $\gamma = 1.0$ curve behaves as an intermediate case, exhibiting a gradual increase in the final steps but still maintaining lower loss in the early stages.

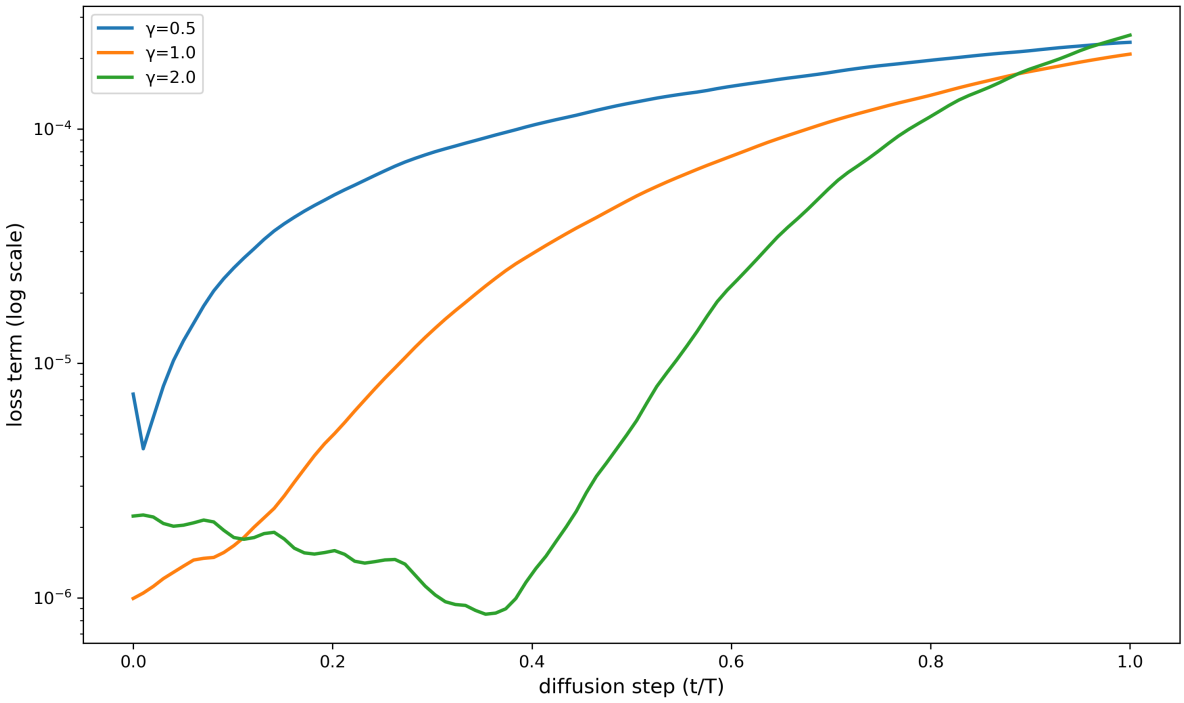


Figure 27 – Loss contribution across diffusion steps for DDPM-100 with different values of γ.

6.2.4 Subset 3mm D45

Tables 22, 23, and 24 report the SSIM, PSNR, and RMSE values for different γ values across the DDPM-10, DDPM-50, and DDPM-100 configurations, respectively, for the 3mm D45 subset. Figures 28, 29 and 30 display the loss contribution per diffusion step for each model.

6.2.4.1 Denoising performance using $T = 10$

In DDPM-10, whose results are shown in Table 22, $\gamma = 2.0$ achieves the best performance, with SSIM=0.8430, PSNR=25.9625, RMSE=20.7183), although the differences between configurations are smaller.

Table 22 – Average SSIM, PSNR, and RMSE metrics on the test set for each γ value tested with DDPM-10.

γ	**SSIM ↑**	**PSNR ↑**	**RMSE ↓**
0.5	0.8362	25.8926	20.8012
1.0	0.8405	25.8318	21.0408
2.0	**0.8430**	**25.9625**	**20.7183**

Figure 28 illustrates the evolution of the loss term across the diffusion steps for each γ configuration in the DDPM-10 setup. The curve for $\gamma = 2.0$ shows significantly lower loss values at earlier steps shows $\gamma = 0.5$ has high initial-step losses, while $\gamma = 1.0$ distributes loss more evenly. This may contribute to the slightly better quantitative results seen in Table 22, particularly in SSIM, because the model is mora specialized in reconstructing fine details. Meanwhile, the curves for $\gamma = 0.5$ and $\gamma = 1.0$ behave extremely similarly, with a more balanced distribution of loss across timesteps.

6.2.4.2 Denoising performance using $T = 50$

As shown in Table 23, the DDPM-50 configuration exhibits relatively small variations in performance across different values of γ. The best overall results are achieved with $\gamma = 2.0$, which yields the highest PSNR (25.6485) and the lowest RMSE (21.4689). However, $\gamma = 0.5$ had a slight higher SSIM value (0.8387).

Figure 29 illustrates the behavior of the loss term across the normalized diffusion steps t/T for different values of γ in the DDPM-50 configuration. the curve for $\gamma = 2.0$ presents a

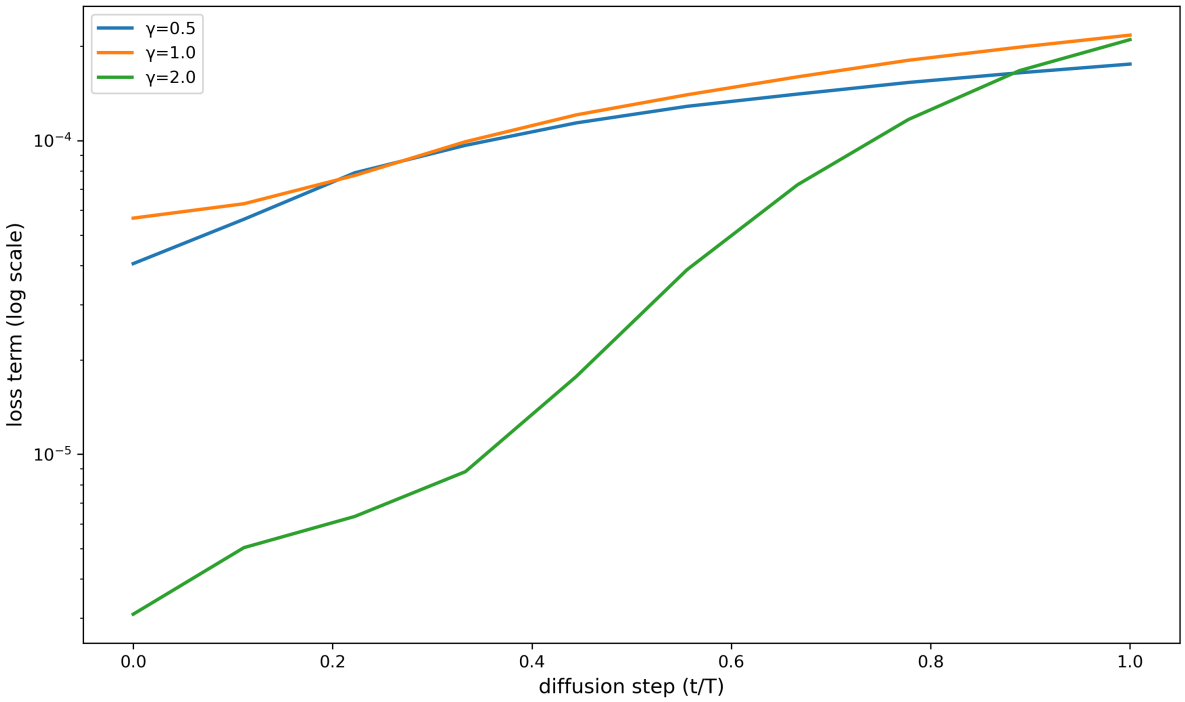


Figure 28 – Loss contribution across diffusion steps for DDPM-10 with different values of γ.

Table 23 – Average SSIM, PSNR, and RMSE metrics on the test set for each γ value tested with DDPM-50.

γ	SSIM ↑	PSNR ↑	RMSE ↓
0.5	**0.8387**	25.0504	23.3512
1.0	0.8353	25.0886	22.7049
2.0	0.8385	**25.6485**	**21.4689**

pronounced U-shape, suggesting that the model emphasized learning at mid-level noise levels. In contrast, for $\gamma = 0.5$, the loss increases steadily throughout the process, with higher loss at final steps. The configuration with $\gamma = 1.0$ shows consistently higher loss values across the entire process.

6.2.4.3 Denoising performance using $T = 100$

Table 24 presents the quantitative results for DDPM-100 across different values of γ. Unlike previous configurations, the results here reveal no consistent trend favoring higher γ values. On the opposite, the best SSIM (0.8403), PSNR (25.6542), and RMSE (21.4031) values were obtained with $\gamma = 0.5$.

Figure 30 illustrates the loss contribution across timesteps for the DDPM-100 model with varying values of γ. When $\gamma = 0.5$, the beginning of the process shows lower losses, which gradually increase as t increases. This early-step focus appears to correlate with the improved

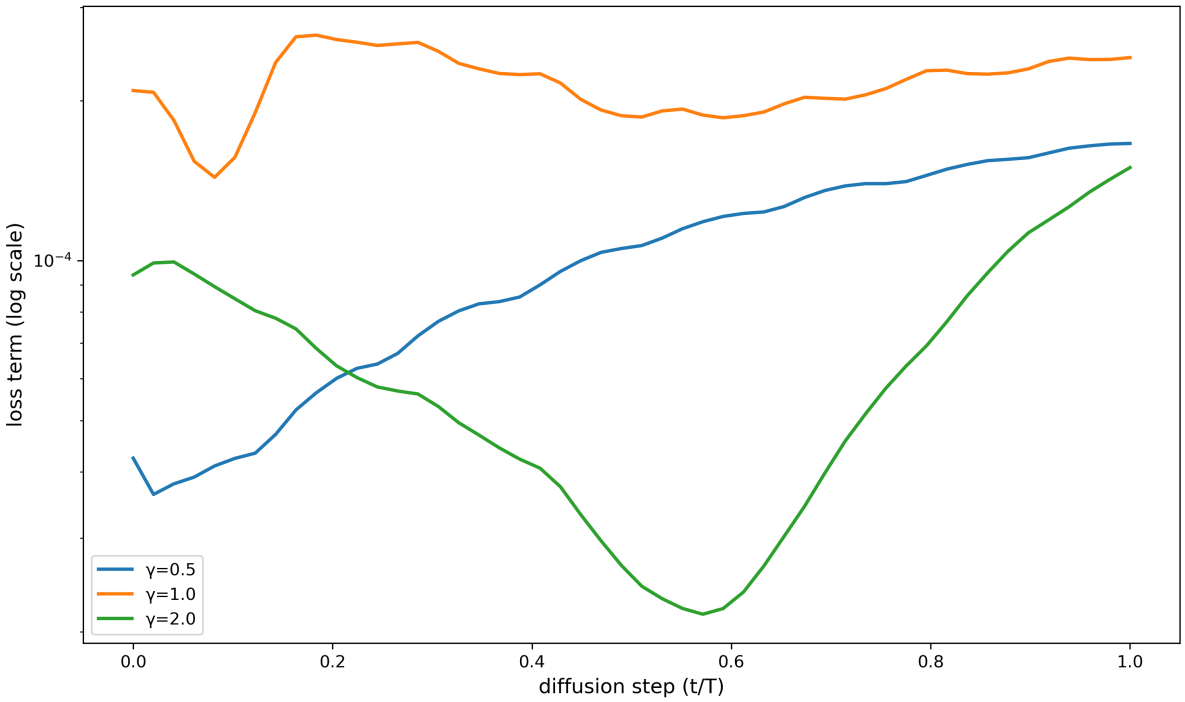


Figure 29 – Loss contribution across diffusion steps for DDPM-50 with different values of γ.

Table 24 – Average SSIM, PSNR, and RMSE metrics on the test set for each γ value tested with DDPM-100.

γ	**SSIM ↑**	**PSNR ↑**	**RMSE ↓**
0.5	**0.8403**	**25.6542**	**21.4031**
1.0	0.8304	25.0537	23.3485
2.0	0.8345	25.3909	22.1356

quantitative results shown in Table 24, especially in terms of SSIM. In contrast, maintains a relatively uniform loss across all steps but yields the weakest performance among the three settings. Surprisingly, the $\gamma = 2.0$ curve behaves much smoother compared to previous presented configurations.

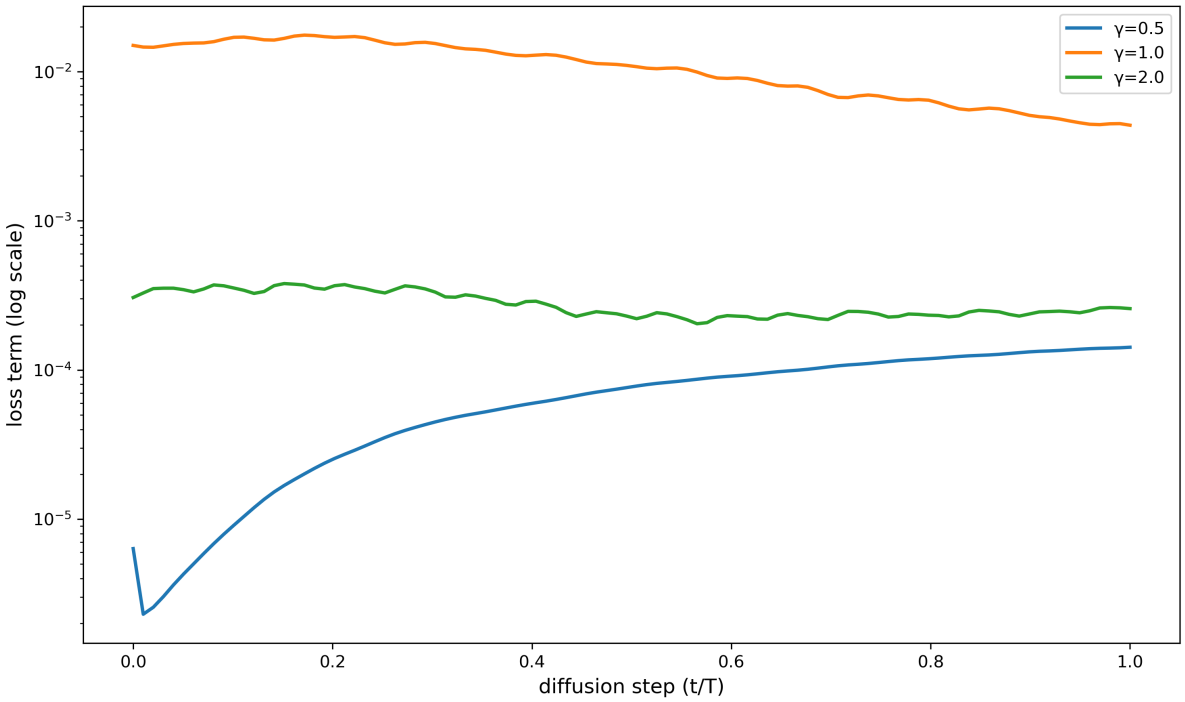


Figure 30 – Loss contribution across diffusion steps for DDPM-100 with different values of γ.

6.3 DISCUSSION

In this section, we focus on discussing the results presented in Sections 6.1 and 6.2. Our findings demonstrate that the diffusion-based models using the proposed power-law noise scheduler consistently outperform the baseline method, which reinforces our hypothesis that iterative denoising leads to more effective denoising than single-step reconstruction techniques. This aligns with recent advances in the literature, where diffusion models have shown superior performance in image restoration tasks, particularly in medical imaging scenarios where high-frequency details are essential (HUNG et al., 2023; WU et al., 2023; WU et al., 2024).

The starting point of our hypothesis was that explicitly modeling the residual noise schedule could modulate the difficulty of the denoising task across the diffusion process. Our experimental results show that $\gamma = 1.0$, which corresponds to the linear schedule, tends to perform worse across all configurations, while the best results were consistently achieved with $\gamma = 2.0$. This validates the superiority of power-lar noise schedules applied to higher-resolution images, since our 512×512 images tend to benefit from more aggressive, noisier schedules that emphasize later denoising steps, as proposed by the works of Nichol e Dhariwal (2021), Chen (2023).

From the analysis of the loss curves, we can already perceive manifestations of the Curriculum Learning effect, since our power-law schedule operationalizes the curriculum strategies as

stated in Section 4.2.1. For instance, for $\gamma = 2.0$, the loss is concentrated at later timesteps, which carry stronger high-frequency structures (KARRAS et al., 2022). The smoother and more stable loss pattern observed for $\gamma = 2.0$ during the initial half of the diffusion process also indicates that the model faces less variability when learning to denoise early steps, which can help the model focus on capturing fine details (QIAN et al., 2024).

For shorter diffusion processes such as $T = 10$, the differences between schedules were less pronounced, even though $\gamma = 2.0$ still achieved the best overall performance. This may be attributed to the small number of diffusion steps, which minimizes the risk of timestep undersampling during training (KINGMA et al., 2021). Similar findings were reported in Gao et al. (2023), Xia et al. (2024), which shows that with fewer steps, the sampling across timesteps becomes inherently more uniform, ensuring that each step contributes more equally to the learning process.

In contrast, the γ parameter trade-off is particularly noticeable in longer schedules such as $T = 100$, where the performance improves consistently as γ increases. This heightened sensitivity to γ may be explained by the fact that higher values of γ concentrate more noise, and consequently more training signal, in the later steps of the forward process. This focus allows the model to better learn from later timesteps, which is often where denoising is easier (ULYANOV; VEDALDI; LEMPITSKY, 2018). Once again our approach aligns with principles from Curriculum Learning (BENGIO et al., 2009), even though our network was trained with random timestep sampling. By modulating the amount of residual noise added, we replicate a learning structure in which earlier or later timesteps are emphasized depending on γ (LIANG; BHARDWAJ; ZHOU, 2024).

Interestingly, lower values of γ, such as 0.5, and even the linear schedule with $\gamma = 1.0$, distribute the noise more uniformly across timesteps. While this may seem more balanced, it can lead to training inconsistency unless the model is trained for a greater number of epochs, as more steps increase the risk of timestep undersampling during training (KIM et al., 2024b).

To better explore the learning potential of each schedule, targeted training strategies can be adopted. For instance, with higher values of γ, where the loss is concentrated in the later timesteps, a warm-up phase could be applied specifically to these steps to accelerate convergence where the learning potential is strongest. On the other hand, for lower values of γ, where the loss is more uniformly spread across the diffusion process, an initial training phase ensuring all timesteps are sampled can prevent the risk of timestep undersampling. Training weighting strategy for different timesteps are important (HANG et al., 2023) because

the diffusion model specializes in high-frequency content at early steps, whereas it generates low-frequency content at later ones (SAMI et al., 2024; QIAN et al., 2024).

This concept is in line with the broader idea of task difficulty-based curriculum learning, as explored by the work of Kim et al. (2024a), which demonstrates that earlier timesteps usually are computationally more challenging and benefit from being introduced progressively during training. Similarly, the idea of structuring goals in a curriculum via diffusion processes has been explored in Sayar et al. (2024). This evoked perspective of the principles of Curriculum Learning is directly reflected in the modulation of noise intensity via the power-law scheduler proposed in our diffusion framework, where different values of γ implicitly define a difficulty progression throughout the diffusion steps.

7 CONCLUSION

In this work, we presented a conditional diffusion-based approach for denoising low-dose computed tomography (LDCT) images, addressing key challenges associated with the traditional way Denoising Diffusion Probabilistic Models address image denoising tasks. Rather than following the standard formulation, which relies on a stochastic Gaussian forward process and requires the network to learn to predict the added noise, we adopt a simplified deterministic forward formulation that has been previously proposed in the literature for image-conditioned diffusion tasks.

This alternative formulation eliminates the need to explicitly define a specific noise distribution during the forward process. As previously discussed, the traditional DDPM pipeline may not be ideally suited for denoising tasks, as the distribution of the synthesized noise may not reflect the actual characteristics of the noise present in real-world data. Instead, the adopted approach directly leverages the noise already present in the noisy image itself, allowing the network to focus on reconstructing the clean image rather than learning to estimate noise.

Building upon this formulation, our main contribution consists in redefining of the noise schedule by introducing a modulation parameter γ, a power-law based modulation factor that controls the intensity of noise addition throughout the diffusion steps. This scheduling mechanism determines how much of the residual noise, which is obtained from the difference between the noisy LDCT image and clean NDCT image, is added at each diffusion step. The power-law γ enables the exploration of different noise schedule routines, which can concentrate noise addition on early or late timesteps.

Our experimental results demonstrate that the proposed method outperforms the baseline, and that the γ parameter plays an important role in performance. In particular, for longer diffusion schedules, higher values of γ lead to better results, suggesting that concentrating noise in the later steps helps the network focus on more relevant parts of the denoising process. We also observed that shorter diffusion processes can produce results comparable to, or even better than, longer ones.

Finally, our analysis reveals that the use of the γ-scheduler introduces an implicit form of task difficulty modulation along the diffusion trajectory, closely related to the principles of Curriculum Learning and Reverse Curriculum Learning. This aligns with recent literature suggesting that controlling the complexity of noise addition over time can benefit training.

7.1 LIMITATIONS

While the proposed method simplifies the forward process and reduces inference time, it is not without limitations. One potential issue is the bias introduced by the deterministic forward process, which may reduce the model's robustness in scenarios with significant variation in noise characteristics. Traditional diffusion models are known for their diversity and stochasticity in sample generation, and although creativity is not relevant for denoising, quality may be compromised in a fully deterministic formulation.

Furthermore, the reliance on interpolation to generate intermediate states during the forward process raises questions about theoretical optimality. While practical, there is still a need for strong theoretical or empirical evidence that this interpolation is the optimal strategy. Additionally, our sampling strategy is adopted directly from the Cold Diffusion framework. Although formally grounded, its suitability as the optimal choice for our specific conditional denoising task needs empirical validation.

Moreover, since the model samples timesteps uniformly during training, some timesteps may be underrepresented, especially in longer schedules. As shown in Figure 31, the comparison of the histograms suggests that, while all timesteps are generally sampled with similar frequency, small variations become more noticeable as the total number of steps increases. This imbalance can affect convergence and denoising performance, particularly in scenarios where the noise schedule distributes noise unevenly throughout the forward process.

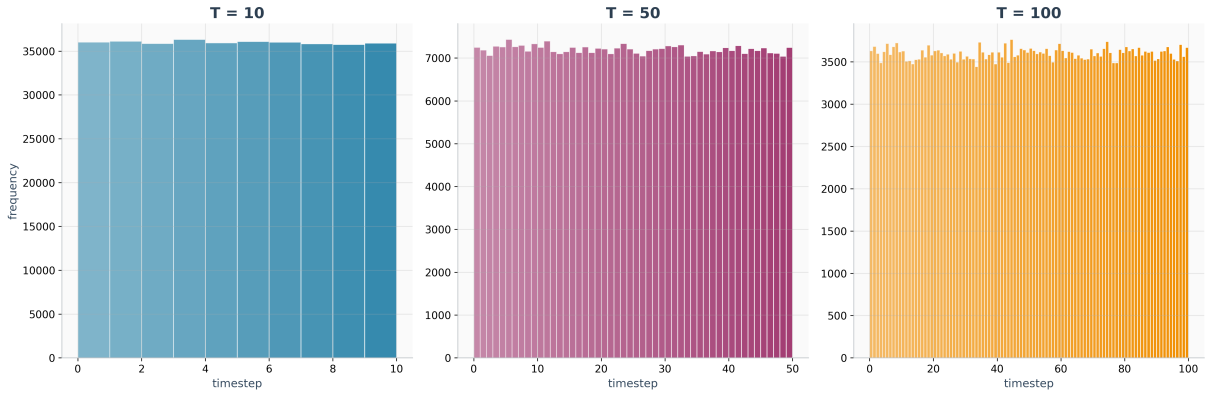


Figure 31 – Histogram comparison of the frequency of sampled timesteps during training for different number of diffusion steps, with $T = 10$, $T = 50$, and $T = 100$, respectively. The model samples timesteps uniformly, but longer schedules may introduce slight imbalances in the frequency of certain timesteps.

Finally, another important limitation of this work is the lack of qualitative evaluation conducted by medical experts. While the quantitative metrics, such as PSNR, SSIM, and RMSE,

provide an objective assessment of the model's performance, they do not fully capture clinical relevance or diagnostic quality. The visual fidelity of reconstructed CT images, especially in detecting subtle anatomical details or pathological structures, must be assessed by radiologists or trained professionals.

7.2 FUTURE RESEARCH

This study opens several possibilities for further research, both in terms of methodological refinements and clinical validation. Future investigations may consider the following directions:

- **Qualitative clinical evaluation:** A crucial step toward real-world applicability involves collaborating with medical professionals to perform qualitative assessments. Specialists could evaluate whether the reconstructions preserve clinically relevant structures and support diagnostic interpretation.
- **Adaptive timestep sampling:** Instead of relying on uniform sampling during training, future models could explore adaptive strategies that prioritize potentially undersampled or more informative timesteps. This could help improve convergence and learning stability, especially in long schedules.
- **Learnable noise schedule (γ):** Instead of setting γ manually, future approaches might benefit from learning the noise scheduling function dynamically during training in response to validation performance.
- **Alternative sampling strategies:** Exploring new or hybrid sampling algorithms, or modifying existing ones could improve the quality of reconstructions.
- **Extension to other imaging modalities and noise types:** Applying the proposed approach to other modalities such as MRI, as well as to more realistic or diverse noise distributions, could test the generalization of the method in different clinical scenarios.
- **Extrapolating residual noise:** An interesting extension would be to extrapolate the use of residual noise, obtained from the difference between noisy and clean images, in the forward process. By expanding this residual noise according to the scheduler, it may be possible to retain the benefits of direct noise prediction while preserving the deterministic formulation.

REFERENCES

AAPM. *Mayo Grand Challenge*. 2016. <https://aapm.app.box.com/s/eaw4jddb53keg1bptavvvd1sf4x3pe9h/file/858370564530>. Accessed: 2025-05-02.

ALMEIDA, M. B. d. *Comparações de modelos neurais em redução de ruído de imagens de tomografia computadorizada*. Dissertação (Dissertação de Mestrado) — Universidade Federal de Pernambuco, Centro de Informática, Recife, 2024. Orientador: Tsang Ing Ren. Inclui referências e anexos.

ARJOVSKY, M.; CHINTALA, S.; BOTTOU, L. Wasserstein generative adversarial networks. In: PRECUP, D.; TEH, Y. W. (Ed.). *Proceedings of the 34th International Conference on Machine Learning*. PMLR, 2017. (Proceedings of Machine Learning Research, v. 70), p. 214–223. Disponível em: <https://proceedings.mlr.press/v70/arjovsky17a.html>.

BANSAL, A.; BORGNIA, E.; CHU, H.-M.; LI, J. S.; KAZEMI, H.; HUANG, F.; GOLDBLUM, M.; GEIPING, J.; GOLDSTEIN, T. *Cold Diffusion: Inverting Arbitrary Image Transforms Without Noise*. 2022. Disponível em: <https://arxiv.org/abs/2208.09392>.

BENGIO, Y.; LOURADOUR, J.; COLLOBERT, R.; WESTON, J. Curriculum learning. In: *Proceedings of the 26th Annual International Conference on Machine Learning (ICML)*. Montreal, Quebec, Canada: ACM, 2009. p. 41–48.

BERRADA, T.; ASTOLFI, P.; HALL, M.; HAVASI, M.; BENCHETRIT, Y.; ROMERO-SORIANO, A.; ALAHARI, K.; DROZDZAL, M.; VERBEEK, J. *Boosting Latent Diffusion with Perceptual Objectives*. 2025. Disponível em: <https://arxiv.org/abs/2411.04873>.

BISHOP, C. M. *Neural Networks for Pattern Recognition*. Oxford, UK: Oxford University Press, 1995. ISBN 978-0198538646.

CHEN, H.; ZHANG, Y.; KALRA, M. K.; LIN, F.; CHEN, Y.; LIAO, P.; ZHOU, J.; WANG, G. Low-dose ct with a residual encoder-decoder convolutional neural network. *IEEE transactions on medical imaging*, IEEE, v. 36, n. 12, p. 2524–2535, 2017.

CHEN, N.; ZHANG, Y.; ZEN, H.; WEISS, R. J.; NOROUZI, M.; CHAN, W. Wavegrad: Estimating gradients for waveform generation. *arXiv preprint arXiv:2009.00713*, 2020.

CHEN, R. T.; BEHRMANN, J.; DUVENAUD, D. K.; JACOBSEN, J.-H. Residual flows for invertible generative modeling. *Advances in Neural Information Processing Systems*, v. 32, 2019.

CHEN, T. *On the Importance of Noise Scheduling for Diffusion Models*. 2023. Disponível em: <https://arxiv.org/abs/2301.10972>.

CHRISTE, A.; HEVERHAGEN, J.; OZDOBA, C.; WEISSTANNER, C.; ULZHEIMER, S.; EBNER, L. Ct dose and image quality in the last three scanner generations. *World journal of radiology*, v. 5, n. 11, p. 421, 2013.

CHUNG, H.; HEO, B.; YOO, J.; KIM, S.; YE, J. C. Come-closer-diffuse-faster: Accelerating conditional diffusion models for inverse problems through stochastic contraction. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. New Orleans, LA, USA: IEEE, 2022. p. 12413–12422.

DHARIWAL, P.; NICHOL, A. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, v. 34, p. 8780–8794, 2021.

DINH, L.; KRUEGER, D.; BENGIO, Y. Nice: Non-linear independent components estimation. *arXiv preprint arXiv:1410.8516*, 2014.

DINH, L.; SOHL-DICKSTEIN, J.; BENGIO, S. Density estimation using real nvp. *arXiv preprint arXiv:1605.08803*, 2016.

EULIG, E.; MCDONAGH, S. G.; CHING, T. W.; CAO, Y.; WANG, G.; RAJCHL, M.; GLOCKER, B.; HAMMERNIK, K. Benchmarking deep learning-based low-dose ct image denoising algorithms. *Medical Physics*, Wiley, v. 51, n. 1, p. 100–115, 2024.

FELLER, W. On the theory of stochastic processes, with particular reference to applications. In: NEYMAN, J. (Ed.). *Proceedings of the First Berkeley Symposium on Mathematical Statistics and Probability*. Berkeley, CA, USA: University of California Press, 1949. v. 1, p. 403–433.

FORD, J. M.; DECKER, S. J. Computed tomography slice thickness and its effects on three-dimensional reconstruction of anatomical structures. *Journal of Forensic Radiology and Imaging*, Elsevier, v. 4, p. 43–46, 2016.

GALLON, D.; JENTZEN, A.; WURSTEMBERGER, P. von. An overview of diffusion models for generative artificial intelligence. *arXiv preprint arXiv:2412.01371*, 2024.

GAO, Q.; LI, Z.; ZHANG, J.; ZHANG, Y.; SHAN, H. Corediff: Contextual error-modulated generalized diffusion model for low-dose ct denoising and generalization. *IEEE Transactions on Medical Imaging*, IEEE, v. 43, n. 2, p. 745–759, 2023.

GOODFELLOW, I. J.; POUGET-ABADIE, J.; MIRZA, M.; XU, B.; WARDE-FARLEY, D.; OZAIR, S.; COURVILLE, A.; BENGIO, Y. *Generative Adversarial Networks*. 2014. Disponível em: <https://arxiv.org/abs/1406.2661>.

GU, J.; ZHAI, S.; ZHANG, Y.; BAUTISTA, M. A.; SUSSKIND, J. f-dm: A multi-stage diffusion model via progressive signal transformation. *arXiv preprint arXiv:2210.04955*, 2022.

GUO, L.; WANG, C.; YANG, W.; HUANG, S.; WANG, Y.; PFISTER, H.; WEN, B. Shadowdiffusion: When degradation prior meets diffusion model for shadow removal. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2023. p. 14049–14058.

GUO, Z.; LANG, J.; HUANG, S.; GAO, Y.; DING, X. A comprehensive review on noise control of diffusion model. *arXiv preprint arXiv:2502.04669*, 2025.

HANG, T.; GU, S.; GENG, X.; GUO, B. Improved noise schedule for diffusion training. *arXiv preprint arXiv:2407.03297*, 2024.

HANG, T.; GU, S.; LI, C.; BAO, J.; CHEN, D.; HU, H.; GENG, X.; GUO, B. Efficient diffusion training via min-snr weighting strategy. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. Paris, France: IEEE, 2023. p. 7441–7451.

HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA: IEEE, 2016. p. 770–778.

HO, J.; CHAN, W.; SAHARIA, C.; FLEET, D.; NOROUZI, M.; SALIMANS, T. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*, 2022.

HO, J.; JAIN, A.; ABBEEL, P. *Denoising Diffusion Probabilistic Models*. 2020. 6840–6851 p. Disponível em: <https://arxiv.org/abs/2006.11239>.

HO, J.; SAHARIA, C.; CHAN, W.; FLEET, D. J.; NOROUZI, M.; SALIMANS, T. Cascaded diffusion models for high fidelity image generation. *Journal of Machine Learning Research*, v. 23, n. 47, p. 1–33, 2022.

HOOGEBOOM, E.; HEEK, J.; SALIMANS, T. Simple diffusion: End-to-end diffusion for high resolution images. In: *Proceedings of the 40th International Conference on Machine Learning (ICML)*. PMLR, 2023. (Proceedings of Machine Learning Research, v. 202), p. 13213–13232. Disponível em: <https://proceedings.mlr.press/v202/hoogeboom23a.html>.

HOU, J.; ZHU, Z.; HOU, J.; LIU, H.; ZENG, H.; YUAN, H. Global structure-aware diffusion process for low-light image enhancement. *Advances in Neural Information Processing Systems*, v. 36, p. 79734–79747, 2023.

HSIEH, J.; NETT, B.; YU, Z.; SAUER, K.; THIBAULT, J.-B.; BOUMAN, C. A. Recent advances in ct image reconstruction. *Current Radiology Reports*, Springer, v. 1, n. 1, p. 39–51, 2013.

HUANG, H.; HE, R.; SUN, Z.; TAN, T. et al. Introvae: Introspective variational autoencoders for photographic image synthesis. *Advances in neural information processing systems*, v. 31, 2018.

HUNG, A. L. Y.; ZHAO, K.; ZHENG, H.; YAN, R.; RAMAN, S. S.; TERZOPOULOS, D.; SUNG, K. Med-cdiff: conditional medical image generation with diffusion models. *Bioengineering*, MDPI, v. 10, n. 11, p. 1258, 2023.

JARZYNSKI, C. Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. *Physical Review E*, American Physical Society (APS), v. 56, n. 5, p. 5018–5035, nov. 1997. ISSN 1095-3787. Disponível em: <http://dx.doi.org/10.1103/PhysRevE.56.5018>.

JARZYNSKI, C. Equalities and inequalities: Irreversibility and the second law of thermodynamics at the nanoscale. In: VACCHINI, B.; ATTAL, S.; BUCHLEITNER, A. (Ed.). *Time: Poincaré Seminar 2010*. Berlin, Heidelberg: Springer, 2012, (Progress in Mathematical Physics, v. 60). p. 145–172.

JOHNSON, J.; ALAHI, A.; FEI-FEI, L. Perceptual losses for real-time style transfer and super-resolution. In: LEIBE, B.; MATAS, J.; SEBE, N.; WELLING, M. (Ed.). *Computer Vision – ECCV 2016*. Cham: Springer, 2016. (Lecture Notes in Computer Science, v. 9906), p. 694–711.

KADKHODAIE, Z.; SIMONCELLI, E. Stochastic solutions for linear inverse problems using the prior implicit in a denoiser. *Advances in Neural Information Processing Systems*, v. 34, p. 13242–13254, 2021.

KALRA, M. K.; PRASAD, S.; SAINI, S.; BLAKE, M. A.; VARGHESE, J.; HALPERN, E. F.; RHEA, J. T.; THRALL, J. H. Clinical comparison of standard-dose and 50% reduced—dose abdominal ct: effect on image quality. *American Journal of Roentgenology*, Am Roentgen Ray Soc, v. 179, n. 5, p. 1101–1106, 2002.

KARRAS, T.; AITTALA, M.; AILA, T.; LAINE, S. Elucidating the design space of diffusion-based generative models. *Advances in neural information processing systems*, v. 35, p. 26565–26577, 2022.

KATARIA, B.; ALTHÉN, J. N.; SMEDBY, Ö.; PERSSON, A.; SÖKJER, H.; SANDBORG, M. Assessment of image quality in abdominal computed tomography: effect of model-based iterative reconstruction, multi-planar reconstruction and slice thickness on potential dose reduction. *European Journal of Radiology*, Elsevier, v. 122, p. 108703, 2020.

KAWAR, B.; ELAD, M.; ERMON, S.; SONG, J. Denoising diffusion restoration models. *Advances in Neural Information Processing Systems*, v. 35, p. 23593–23606, 2022.

KAWAR, B.; VAKSMAN, G.; ELAD, M. Stochastic image denoising by sampling from the posterior distribution. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. Montreal, QC, Canada: IEEE, 2021. p. 1866–1875.

KIM, J.-Y.; GO, H.; KWON, S.; KIM, H.-G. Denoising task difficulty-based curriculum for training diffusion models. *arXiv preprint arXiv:2403.10348*, 2024.

KIM, M.; KI, D.; SHIM, S.-W.; LEE, B.-J. *Adaptive Non-Uniform Timestep Sampling for Diffusion Model Training*. 2024. Disponível em: <https://arxiv.org/abs/2411.09998>.

KINGMA, D.; SALIMANS, T.; POOLE, B.; HO, J. Variational diffusion models. *Advances in neural information processing systems*, v. 34, p. 21696–21707, 2021.

KINGMA, D. P.; BA, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

KINGMA, D. P.; DHARIWAL, P. Glow: Generative flow with invertible 1x1 convolutions. *Advances in neural information processing systems*, v. 31, 2018.

KINGMA, D. P.; SALIMANS, T.; JOZEFOWICZ, R.; CHEN, X.; SUTSKEVER, I.; WELLING, M. Improved variational inference with inverse autoregressive flow. *Advances in neural information processing systems*, v. 29, 2016.

KINGMA, D. P.; WELLING, M. *Auto-Encoding Variational Bayes*. 2013. Disponível em: <https://arxiv.org/abs/1312.6114>.

KISTLER, J. P.; HOCHBERG, F. H.; BROOKS, B. R.; JR, E. R.; NEW, P.; SCHNUR, J. Computerized axial tomography: clinicopathologic correlation. *Neurology*, AAN Enterprises, v. 25, n. 3, p. 201–201, 1975.

KREIS, K.; GAO, R.; VAHDAT, A. Denoising diffusion-based generative modeling: Foundations and applications. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. New Orleans, LA, USA: IEEE, 2022.

KRILLE, L.; HAMMER, G. P.; MERZENICH, H.; ZEEB, H. Systematic review on physician's knowledge about radiation doses and radiation risks of computed tomography. *European Journal of Radiology*, v. 76, n. 1, p. 36–41, 2010. ISSN 0720-048X. Radiation Protection in Medical Imaging - Never Ending Story. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0720048X10004018>.

KULATHILAKE, K. S. H.; ABDULLAH, N. A.; SABRI, A. Q. M.; LAI, K. W. A review on deep learning approaches for low-dose computed tomography restoration. *Complex & Intelligent Systems*, Springer, v. 9, n. 3, p. 2713–2745, 2023.

LI, M.; HSU, W.; XIE, X.; CONG, J.; GAO, W. Sacnn: Self-attention convolutional neural network for low-dose ct denoising with self-supervised perceptual loss network. *IEEE transactions on medical imaging*, IEEE, v. 39, n. 7, p. 2289–2301, 2020.

LIANG, Y.; BHARDWAJ, S.; ZHOU, T. Diffusion curriculum: Synthetic-to-real generative curriculum learning via image-guided diffusion. *arXiv preprint arXiv:2410.13674*, 2024.

LIN, S.; YANG, X. *Diffusion Model with Perceptual Loss*. 2025. Disponível em: <https://arxiv.org/abs/2401.00110>.

LIU, L. Model-based iterative reconstruction: a promising algorithm for today's computed tomography imaging. *Journal of Medical imaging and Radiation sciences*, Elsevier, v. 45, n. 2, p. 131–136, 2014.

LIU, X.; XIE, Y.; LIU, C.; CHENG, J.; DIAO, S.; TAN, S.; LIANG, X. Diffusion probabilistic priors for zero-shot low-dose ct image denoising. *Medical Physics*, Wiley Online Library, v. 52, n. 1, p. 329–345, 2023.

LUO, C. Understanding diffusion models: A unified perspective. *arXiv preprint arXiv:2208.11970*, 2022.

MA, Z.; ZHANG, Y.; JIA, G.; ZHAO, L.; MA, Y.; MA, M.; LIU, G.; ZHANG, K.; DING, N.; LI, J. et al. Efficient diffusion models: A comprehensive survey from principles to practices. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE, 2025.

Mayo Clinic. *2016 Low Dose CT Grand Challenge*. 2016. <https://ctcicblog.mayo.edu/2016-low-dose-ct-grand-challenge/>. Accessed: 2025-05-02.

MCCOLLOUGH, C. H.; BARTLEY, A. C.; CARTER, R. E.; CHEN, B.; DREES, T. A.; EDWARDS, P.; III, D. R. H.; HUANG, A. E.; KHAN, F.; LENG, S. et al. Low-dose ct for the detection and classification of metastatic liver lesions: results of the 2016 low dose ct grand challenge. *Medical physics*, Wiley Online Library, v. 44, n. 10, p. e339–e352, 2017.

MENG, C.; HE, Y.; SONG, Y.; SONG, J.; WU, J.; ZHU, J.-Y.; ERMON, S. Sdedit: Guided image synthesis and editing with stochastic differential equations. *arXiv preprint arXiv:2108.01073*, 2022. Disponível em: <https://arxiv.org/abs/2108.01073>.

METTLER, F. A.; WIEST, P. W.; LOCKEN, J. A.; KELSEY, C. A. Ct scanning: patterns of use and dose. *Journal of radiological Protection*, IOP Publishing, v. 20, n. 4, p. 353, 2000.

MEULEPAS, J. M.; RONCKERS, C. M.; SMETS, A. M.; NIEVELSTEIN, R. A.; GRADOWSKA, P.; LEE, C.; JAHNEN, A.; STRATEN, M. van; WIT, M.-C. Y. de; ZONNENBERG, B. et al. Radiation exposure from pediatric ct scans and subsequent cancer risk in the netherlands. *JNCI: Journal of the National Cancer Institute*, Oxford University Press, v. 111, n. 3, p. 256–263, 2019.

MURPHY, A.; MCKAY, L.; FEGER, J. et al. *Iterative reconstruction (CT)*. 2025. Radiopaedia.org Reference Article. Accessed on 28 Oct 2025. Disponível em: <https://radiopaedia.org/articles/iterative-reconstruction-ct?lang=us>.

NEAL, R. M. *Annealed Importance Sampling*. 1998. Disponível em: <https://arxiv.org/abs/physics/9803008>.

NICHOL, A.; DHARIWAL, P. Improved denoising diffusion probabilistic models. In: *Proceedings of the 38th International Conference on Machine Learning (ICML)*. PMLR, 2021. (Proceedings of Machine Learning Research, v. 139), p. 8162–8171. Disponível em: <http://proceedings.mlr.press/v139/nichol21a.html>.

NICHOL, A.; DHARIWAL, P.; RAMESH, A.; SHYAM, P.; MISHKIN, P.; MCGREW, B.; SUTSKEVER, I.; CHEN, M. Glide: Towards photorealistic image generation and editing with text-guided diffusion models. *arXiv preprint arXiv:2112.10741*, 2021.

ÖZDENIZCI, O.; LEGENSTEIN, R. Restoring vision in adverse weather conditions with patch-based denoising diffusion models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE, v. 45, n. 8, p. 10346–10357, 2023.

PASZKE, A.; GROSS, S.; MASSA, F.; LERER, A.; BRADBURY, J.; CHANAN, G.; KILLEEN, T.; LIN, Z.; GIMELSHEIN, N.; ANTIGA, L.; DESMAISON, A.; KOPF, A.; YANG, E.; DEVITO, Z.; RAISON, M.; TEJANI, A.; CHILAMKURTHY, S.; STEINER, B.; FANG, L.; BAI, J.; CHINTALA, S. Pytorch: An imperative style, high-performance deep learning library. In: WALLACH, H.; LAROCHELLE, H.; BEYGELZIMER, A.; ALCHé-BUC, F. d'; FOX, E.; GARNETT, R. (Ed.). *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2019. v. 32. Disponível em: <https://proceedings.neurips.cc/paper_files/paper/2019/file/bdbca288fee7f92f2bfa9f7012727740-Paper.pdf>.

PATEL, P. R.; JESUS, O. D. *CT Scan*. StatPearls Publishing, Treasure Island (FL), 2023. Disponível em: <http://europepmc.org/books/NBK567796>.

PELT, D. M.; BATENBURG, K. J. Fast tomographic reconstruction from limited data using artificial neural networks. *IEEE Transactions on Image Processing*, v. 22, n. 12, p. 5238–5251, 2013.

QIAN, Y.; CAI, Q.; PAN, Y.; LI, Y.; YAO, T.; SUN, Q.; MEI, T. Boosting diffusion models with moving average sampling in frequency domain. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Vancouver, BC, Canada: IEEE, 2024. p. 8911–8920.

QIN, C.; SCHLEMPER, J.; CABALLERO, J.; PRICE, A. N.; HAJNAL, J. V.; RUECKERT, D. Convolutional recurrent neural networks for dynamic mr image reconstruction. *IEEE Transactions on Medical Imaging*, v. 38, n. 1, p. 280–290, 2019.

RAMESH, A.; DHARIWAL, P.; NICHOL, A.; CHU, C.; CHEN, M. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*, v. 1, n. 2, p. 3, 2022.

RAMESH, A.; PAVLOV, M.; GOH, G.; GRAY, S.; VOSS, C.; RADFORD, A.; CHEN, M.; SUTSKEVER, I. Zero-shot text-to-image generation. In: *Proceedings of the 38th International Conference on Machine Learning (ICML)*. PMLR, 2021. (Proceedings of Machine Learning Research, v. 139), p. 8821–8831. Disponível em: <http://proceedings.mlr.press/v139/ramesh21a.html>.

RAMPINELLI, C.; ORIGGI, D.; BELLOMI, M. Low-dose ct: technique, reading methods and image interpretation. *Cancer imaging*, v. 12, n. 3, p. 548, 2013.

ROMBACH, R.; BLATTMANN, A.; LORENZ, D.; ESSER, P.; OMMER, B. *High-Resolution Image Synthesis with Latent Diffusion Models*. 2022. Disponível em: <https://arxiv.org/abs/2112.10752>.

RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: NAVAB, N.; HORNEGGER, J.; WELLS, W. M.; FRANGI, A. F. (Ed.). *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Cham: Springer, 2015. (Lecture Notes in Computer Science, v. 9351), p. 234–241.

RUTHOTTO, L.; HABER, E. An introduction to deep generative modeling. *GAMM-Mitteilungen*, Wiley Online Library, v. 44, n. 2, p. e202100008, 2021.

SAHARIA, C.; CHAN, W.; CHANG, H.; LEE, C.; HO, J.; SALIMANS, T.; FLEET, D.; NOROUZI, M. Palette: Image-to-image diffusion models. In: *ACM SIGGRAPH 2022 Conference Proceedings*. New York, NY, USA: Association for Computing Machinery, 2022. (SIGGRAPH '22). ISBN 9781450393379. Disponível em: <https://doi.org/10.1145/3528233.3530757>.

SAHARIA, C.; CHAN, W.; SAXENA, S.; LI, L.; WHANG, J.; DENTON, E. L.; GHASEMIPOUR, K.; LOPES, R. G.; AYAN, B. K.; SALIMANS, T. et al. Photorealistic text-to-image diffusion models with deep language understanding. *Advances in neural information processing systems*, v. 35, p. 36479–36494, 2022.

SAHARIA, C.; HO, J.; CHAN, W.; SALIMANS, T.; FLEET, D. J.; NOROUZI, M. Image super-resolution via iterative refinement. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 45, n. 4, p. 4713–4726, 2022.

SAHOO, S.; GOKASLAN, A.; SA, C. M. D.; KULESHOV, V. Diffusion models with learned adaptive noise. *Advances in Neural Information Processing Systems*, v. 37, p. 105730–105779, 2024.

SALIMANS, T.; KARPATHY, A.; CHEN, X.; KINGMA, D. P. Pixelcnn++: Improving the pixelcnn with discretized logistic mixture likelihood and other modifications. *arXiv preprint arXiv:1701.05517*, 2017.

SAMI, S. M.; HASAN, M. M.; DAWSON, J.; NASRABADI, N. Hf-diff: High-frequency perceptual loss and distribution matching for one-step diffusion-based image super-resolution. *arXiv e-prints*, p. arXiv–2411, 2024.

SAYAR, E.; IACCA, G.; OGUZ, O. S.; KNOLL, A. Diffusion-based curriculum reinforcement learning. *Advances in Neural Information Processing Systems*, v. 37, p. 97587–97617, 2024.

SHI, Z.; XU, C.; DONG, C.; PAN, B.; HE, A.; LI, T.; FU, H. et al. Resfusion: Denoising diffusion probabilistic models for image restoration based on prior residual noise. *Advances in Neural Information Processing Systems*, v. 37, p. 130664–130693, 2024.

SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

SOHL-DICKSTEIN, J.; WEISS, E. A.; MAHESWARANATHAN, N.; GANGULI, S. *Deep Unsupervised Learning using Nonequilibrium Thermodynamics*. 2015. 2256–2265 p. Disponível em: <https://arxiv.org/abs/1503.03585>.

SONG, J.; MENG, C.; ERMON, S. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020.

STRüMKE, I.; LANGSETH, H. *Lecture Notes in Probabilistic Diffusion Models*. 2023. Disponível em: <https://arxiv.org/abs/2312.10393>.

THANH-TUNG, H.; TRAN, T. Catastrophic forgetting and mode collapse in gans. In: *2020 International Joint Conference on Neural Networks (IJCNN)*. Glasgow, UK: IEEE, 2020. p. 1–10.

THIBAULT, J.-B.; SAUER, K. D.; BOUMAN, C. A.; HSIEH, J. A three-dimensional statistical approach to improved image quality for multislice helical ct. *Medical physics*, Wiley Online Library, v. 34, n. 11, p. 4526–4544, 2007.

ULYANOV, D.; VEDALDI, A.; LEMPITSKY, V. Deep image prior. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Salt Lake City, UT, USA: IEEE, 2018. p. 9446–9454.

VASWANI, A.; SHAZEER, N.; PARMAR, N.; USZKOREIT, J.; JONES, L.; GOMEZ, A. N.; KAISER, L. u.; POLOSUKHIN, I. Attention is all you need. In: GUYON, I.; LUXBURG, U. V.; BENGIO, S.; WALLACH, H.; FERGUS, R.; VISHWANATHAN, S.; GARNETT, R. (Ed.). *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2017. v. 30. Disponível em: <https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf>.

WANG, T.; ZHANG, K.; ZHANG, Y.; LUO, W.; STENGER, B.; LU, T.; KIM, T.-K.; LIU, W. Lldiffusion: Learning degradation representations in diffusion models for low-light image enhancement. *Pattern Recognition*, Elsevier, p. 111628, 2025.

WANG, Z.; BOVIK, A. C. Mean squared error: Love it or leave it? a new look at signal fidelity measures. *IEEE signal processing magazine*, IEEE, v. 26, n. 1, p. 98–117, 2009.

WANG, Z.; BOVIK, A. C.; SHEIKH, H. R.; SIMONCELLI, E. P. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, IEEE, v. 13, n. 4, p. 600–612, 2004.

WEINSHALL, D.; COHEN, G.; AMIR, D. Curriculum learning by transfer learning: Theory and experiments with deep networks. In: *Proceedings of the 35th International Conference on Machine Learning (ICML)*. PMLR, 2018. (Proceedings of Machine Learning Research, v. 80), p. 5238–5246. Disponível em: <http://proceedings.mlr.press/v80/weinshall18a.html>.

WENG, L. What are diffusion models? *lilianweng.github.io*, Jul 2021. Disponível em: <https://lilianweng.github.io/posts/2021-07-11-diffusion-models/>.

WHANG, J.; DELBRACIO, M.; TALEBI, H.; SAHARIA, C.; DIMAKIS, A. G.; MILANFAR, P. Deblurring via stochastic refinement. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. New Orleans, LA, USA: IEEE, 2022. p. 16293–16303.

WOLTERINK, J. M.; LEINER, T.; VIERGEVER, M. A.; IŠGUM, I. Generative adversarial networks for noise reduction in low-dose ct. *IEEE transactions on medical imaging*, IEEE, v. 36, n. 12, p. 2536–2545, 2017.

WU, J.; FU, R.; FANG, H.; ZHANG, Y.; YANG, Y.; XIONG, H.; LIU, H.; XU, Y. Medsegdiff: Medical image segmentation with diffusion probabilistic model. In: *Medical Imaging with Deep Learning*. [s.n.], 2023. Disponível em: <https://openreview.net/forum?id=Jdw-cm2jG9>.

WU, J.; JI, W.; FU, H.; XU, M.; JIN, Y.; XU, Y. Medsegdiff-v2: Diffusion-based medical image segmentation with transformer. *Proceedings of the AAAI Conference on Artificial Intelligence*, v. 38, n. 6, p. 6030–6038, Mar. 2024. Disponível em: <https://ojs.aaai.org/index.php/AAAI/article/view/28418>.

WU, Q.; GAO, R.; ZHA, H. Bridging explicit and implicit deep generative models via neural stein estimators. *Advances in Neural Information Processing Systems*, v. 34, p. 11274–11286, 2021.

WU, Y.; HE, K. Group normalization. In: *Computer Vision – ECCV 2018: 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part XIII*. Berlin, Heidelberg: Springer-Verlag, 2018. p. 3–19. ISBN 978-3-030-01260-1. Disponível em: <https://doi.org/10.1007/978-3-030-01261-8_1>.

XIA, M.; ZHOU, Y.; YI, R.; LIU, Y.-J.; WANG, W. A diffusion model translator for efficient image-to-image translation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE, 2024.

XIA, W.; LYU, Q.; WANG, G. *Low-Dose CT Using Denoising Diffusion Probabilistic Model for 20× Speedup*. 2022. Disponível em: <https://arxiv.org/abs/2209.15136>.

XIE, Y.; YUAN, M.; DONG, B.; LI, Q. *Diffusion Model for Generative Image Denoising*. 2023. Disponível em: <https://arxiv.org/abs/2302.02398>.

XU, K.; LU, S.; HUANG, B.; WU, W.; LIU, Q. Stage-by-stage wavelet optimization refinement diffusion model for sparse-view ct reconstruction. *IEEE Transactions on Medical Imaging*, IEEE, v. 43, n. 10, p. 3412–3424, 2024.

YANG, Q.; YAN, P.; ZHANG, Y.; YU, H.; SHI, Y.; MOU, X.; KALRA, M. K.; ZHANG, Y.; SUN, L.; WANG, G. Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss. *IEEE Transactions on Medical Imaging*, v. 37, n. 6, p. 1348–1357, 2018.

YU, S.; DONG, H.; YANG, G.; SLABAUGH, G.; DRAGOTTI, P. L.; YE, X.; LIU, F.; ARRIDGE, S.; KEEGAN, J.; FIRMIN, D. et al. Deep de-aliasing for fast compressive sensing mri. *arXiv preprint arXiv:1705.07137*, 2017.

ZENG, R.; LIN, C. Y.; LI, Q.; JIANG, L.; SKOPEC, M.; FESSLER, J. A.; MYERS, K. J. Performance of a deep learning-based ct image denoising method: Generalizability over dose, reconstruction kernel, and slice thickness. *Medical physics*, Wiley Online Library, v. 49, n. 2, p. 836–853, 2022.

ZHANG, M.; GU, S.; SHI, Y. The use of deep learning methods in low-dose computed tomography image reconstruction: a systematic review. *Complex & intelligent systems*, Springer, v. 8, n. 6, p. 5545–5561, 2022.