ENHANCING HISTOPATHOLOGY IMAGE GENERATION WITH DIFFUSION GENERATIVE MODELS: A COMPREHENSIVE STUDY

Denisha Thakkar

A Thesis

in

The Department

of

Computer Science and Software Engineering

Presented in Partial Fulfillment of the Requirements

for the Degree of

Master of Science (Computer Science) at

Concordia University

Montréal, Québec, Canada

September 2024

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: **Denisha Thakkar**

Entitled: **ENHANCING HISTOPATHOLOGY IMAGE GENERATION WITH DIFFUSION GENERATIVE MODELS: A COMPREHENSIVE STUDY**

and submitted in partial fulfillment of the requirements for the degree of

Master of Science (Computer Science)

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the Final Examining Committee:

_____ Chair
Dr. Eugene Belilovsky

_____ Examiner
Dr. Hassan Rivaz

_____ Examiner
Dr. Eugene Belilovsky

_____ Supervisor
Dr. Mahdi S. Hossieni

Approved by _____
Joey Paquet, Chair
Department of Computer Science and Software Engineering

_____ 2024

_____
Mourad Debbabi, Dean
Faculty of Engineering and Computer Science

Abstract

ENHANCING HISTOPATHOLOGY IMAGE GENERATION WITH DIFFUSION GENERATIVE MODELS: A COMPREHENSIVE STUDY

Denisha Thakkar

The field of histopathology faces significant challenges due to the limited availability of data, which is often not publicly accessible due to privacy issues. The scarcity of high-quality publicly available datasets hampers the development and training of effective deep learning models. Generative Adversarial Networks (GANs) have previously attempted to address these issues by creating synthetic data but suffer mode collapse, which reduces their effectiveness and reliability. This study explores Diffusion Generative Models (DGMs) as a unique and robust alternative for generating synthetic pathology images.

The primary objective of this study is to compare various Diffusion Generative Models (DGMs) and methods in medical imaging. Specifically, we examine the Denoising Diffusion Probabilistic Model (DDPM) and the Latent Diffusion Model (LDM), along with other generative sampling choices. Both models demonstrated the ability to generate realistic histopathological images. We also investigated DGMs from a unique perspective by generating various patch sizes, demonstrating that DGMs effectively learn patch resolution.

We analyzed the impact of DGMs on different magnifications in the KGH dataset, focusing on image patches of 224x224 and 336x336 pixels. Larger patches (336x336) showed better performance, with real data achieving 94.06% accuracy and generated data 92.44%. However, combining real and generated data slightly reduced accuracy to 90.76%. For 224x224 patches, real data achieved 89.95%, generated data 88.62%, and combined data improved to 90.75%. These results indicate that synthetic data enhances model performance, particularly with larger image patches.

In computational pathology, generative models can enhance data sharing and augmentation, improving the accuracy of deep learning classifiers, and assisting in the cancer diagnosis workflow,

thereby advancing digital pathology. Our research findings confirm this and set the stage for future developments in the pathology workflow.

Acknowledgments

I want to express my utmost gratitude to my supervisor, Dr. Mahdi Hosseini, for his constant support and guidance throughout my project. His expertise, valuable insights, and feedback have been instrumental in shaping the direction of my research and enhancing the quality of my work. I would also like to acknowledge the contributions of my colleagues, Cassandre and Vasudev, whose work has greatly assisted me in my endeavors. Additionally, I extend my thanks to Huron Pathology for their support in providing the pathology dataset.

I am deeply grateful to my previous friend, Kevin Jivani, whose input and knowledge in the deep learning domain greatly helped me start this research journey and discuss ideas. I am also immensely thankful to my family, my husband, Hariom Sharma, without whom I could not have achieved this. His unwavering support, and belief in me have been the driving force behind all this work. I want to also thank my parents Reena and Mahesh Thakkar, and my siblings, Diya and Prem, for their boundless sacrifices and tireless efforts. Their encouragement has been the pillar of my success, and I am truly indebted to their contributions.

Lastly, I sincerely thank Concordia University for giving me this platform.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 General Background and Motivation

Histopathology involves diagnosing diseases by closely inspecting gigapixel tissue slides of microscopic structures to identify their characteristics (Jahn, Plass, & Moinfar, 2020). Computational pathology has gained significant momentum, bringing about a transformative change in the field of cancer diagnostics after the digitization of pathology slides (Baxi, Edwards, Montalto, & Saha, 2022). In addition, the success of deep learning methods (Janowczyk & Madabhushi, 2016) has led to the development of numerous models that enhance histopathology diagnosis and can also assist pathologists in their diagnosis workflow (Echle et al., 2021; Van der Laak, Litjens, & Ciompi, 2021). However, these models require large volumes of both annotated and unannotated data for effective analysis and cancer screening solutions (Cho, Lee, Shin, Choy, & Do, 2015).

Two primary challenges, related to data issues, influence the development and implementation of computational pathology foundation models in clinical settings. First, privacy concerns severely limit the availability of data, making many pathological data inaccessible to the public (Price & Cohen, 2019). Patient confidentiality and strict regulatory requirements often prevent the sharing of medical data, which is crucial for training and validating deep learning models. This lack of accessible data limits the ability to create the most-optimized and generalizable models. Therefore, without access to diverse data for research, it becomes challenging to develop models that are truly representative and effective across various patient populations and conditions.

Second, the limited quantity and variable quality of publicly available pathology data significantly undermine the performance of foundation models (Daniel et al., 2023). Public datasets often lack the breadth and depth needed to train comprehensive models. The available data can be inconsistent in terms of resolution, staining techniques, and annotation quality (Hosseini et al., 2024). This issue is particularly pronounced for rare cancers, where there is a shortage of various examples for training. Without sufficient high-quality data, models may struggle to accurately diagnose and predict outcomes for these less common conditions.

Another major challenge with pathology data is its complexity, which requires experienced pathologists to analyze and interpret it. The medical field is always looking for new patterns in cancer tissues to diagnose and treat diseases early (O'connor et al., 2017). Digital pathology improves the efficiency of diagnostic procedures by effectively capturing, sharing, and interpreting pathological specimens using digital technology (Kiran et al., 2023). It deals with gigapixel images that have multiple magnification levels and unique features (Kuklyte et al., 2021), which are different from typical computer vision data. Studying these images and creating synthetic datasets requires a deep understanding of their unique traits. This knowledge is essential to accurately replicate pathological data in the synthesis process, helping to advance data processing and research. These challenges highlight the need for innovative solutions to improve data sharing and improve the quality of publicly available pathology datasets. Addressing these issues is essential for advancing the field of computational pathology and improving diagnostic workflow capabilities in clinical settings.

With traditional deep learning applications, such as image classification, high-dimensional image data is distilled into single-class labels. In contrast, the generation of images from class labels is an emerging and promising area (Han, Zheng, & Zhou, 2022; Qin, Zheng, Yao, Zhou, & Zhang, 2023), highlighting the vast potential of generative models to revolutionize pathology. Following the work of (Dhariwal & Nichol, 2021), diffusion models have been improved with conditioning capabilities, significantly increasing their flexibility (Carlini et al., 2023; Rombach, Blattmann, Lorenz, Esser, & Ommer, 2022). We were motivated to study these models in pathology due to its numerous potential usecases in education, clinical quality assurance, improving deep learning classifiers, and digital image processing. This flexibility allows us to generate more images of a particular class,

which can help address the common issue of class imbalance where some classes are underrepresented, and can impact the performance of deep learning models. By generating more images for these underrepresented classes, we can create a more balanced dataset, improving the training and performance of classifiers (Oh & Jeong, 2023).

Generative models in pathology are still in their early stages compared to classification models. However, they hold immense potential, especially in the realm of synthetic medical image generation, which can help address the challenges mentioned above in this field. This approach is particularly promising for overcoming privacy concerns as mentioned , as synthetic data maintain statistical properties similar to real data without compromising patient confidentiality (Savage, 2023). This significantly expands the potential for the use of more data in research and education (Levine et al., 2020). Moreover, by creating synthetic images, generative models can become helpful in discovering new patterns and regularities that provide deeper insights for rare tissue types (Chen, Lu, Chen, Williamson, & Mahmood, 2021).

In particular of Generative models, Generative Adversarial Networks (GANs) (Goodfellow et al., 2020) have set a high standard in creating high-fidelity and quality data patterns, greatly benefiting fields that demand high realism, such as medical imaging (Falahkheirkhah et al., 2023; Kapil et al., 2018; Quiros, Murray-Smith, & Yuan, 2019; Y. Tang, Tang, Zhu, Xiao, & Summers, 2021; Zhou, Fu, Chen, Shen, & Shao, 2020). Despite GANs' achievements, their real application has many challenges, such as mode collapse and training instability, which are critical setbacks in sensitive applications such as medical imaging. Recently, diffusion generative models (DGMs) (Dhariwal & Nichol, 2021; Ho, Jain, & Abbeel, 2020; Nichol & Dhariwal, 2021; Y. Song & Ermon, 2019, 2020; Y. Song et al., 2020) have emerged with superior image generation capabilities and great model stability. To date, diffusion models have been found to be useful in a wide variety of areas, ranging from generative modeling tasks such as image generation (Dhariwal & Nichol, 2021), image super-resolution (H. Li et al., 2022) to discriminative tasks such as image segmentation (Oh & Jeong, 2023), classification (Han et al., 2022). Diffusion Generative Models(DGMs) offer a structured approach based on the strong mathematical foundation, which makes them more reliable to use and offers great flexibility in terms of image generation.

Although Diffusion Generative Models (DGMs) have been applied in various fields for multiple

tasks, such as text-conditioned generation (Yellapragada et al., 2024), image generation (Moghadam et al., 2023), and large-scale image generation (Graikos et al., 2024), their potential for generating synthetic datasets and their applicability in tissue classification are still less explored. Histopathological images are unique in their complexity and detail, which varies significantly between different magnification levels or patch sizes. Effectively generating synthetic images that accurately represent these different levels of detail is a critical, yet under-researched area. In addition, there is a lack of comprehensive comparisons and recommendations on how a complete DGM framework and its associated methods could be effectively used for pathology datasets in conjunction with real data. This integration is crucial because combining synthetic and real data can enhance model training, providing more robust and generalizable results.

In our research, we have addressed this gap using multiple subset datasets from the same private Kingston General Hospital (KGH) data. By doing so, we were able to compare the differences in the synthesizing various fields of view (FOV) (Basavanhally et al., 2011) from the same whole slides, specifically focusing on FOVs of 224 and 336. Different regions of a histopathology slide may require different levels of detail, from high-level overviews to detailed cellular structures. Our approach aims to highlight both the potential and the challenges of applying DGMs to generate these diverse synthetic images. We also provide a novel study on the effects of prompting patch size, which further helps to understand how different patch sizes can influence FOV detail generation. This study shows that by prompting with different sizes, DGMs can generate images with varying levels of detail that are unseen during training.

In summary, our objective was to leverage advancements in diffusion generative models (DGM) to address critical data challenges in computational pathology. We focused on generating high-quality synthetic pathology images to mitigate issues related to data privacy and the scarcity of annotated data. Using subsets of the Kingston General Hospital (KGH) dataset, we compared the synthesis of field of view (FOV) of 224 and 336. This comparison provided insights into the effectiveness of DGMs in replicating different levels of detail in histopathological images and generating high-quality synthetic datasets. These synthetic datasets can significantly improve the robustness and accuracy of deep learning classifiers as well.

1.2 Contributions

This thesis includes a significant body of work that includes a comprehensive literature review, the main body of a soon to be published paper in the *Medical Image Analysis* journal, and additional results and analysis. The study represents a novel exploration of diffusion generative models (DGMs) in the field of computational pathology. It offers a comprehensive comparison of various baseline methods and showcases the ability of diffusion models to create coarse features and generate images with varying patch size.

- **Literature Review:** It focuses on the initial review of Diffusion Generative Models in Chapter 2. In this chapter, various methods are identified that are already utilized in the Computer Vision domain for various tasks, were thoroughly reviewed.

- **Comparison of FOVs:** In our investigation, we have addressed critical gaps utilizing multiple subset datasets from the private Kingston General Hospital (KGH) data. By synthesizing various fields of view (FOV) from the same dataset, specifically focusing on FOVs of 224 and 336, we have been able to compare the differences in FID score by diffusion-generative models (DGM).

- **Comprehensive Analysis:** This thesis includes a comprehensive study in Section 4.2.1 that compares various diffusion generative methods. The study provides a detailed analysis of the strengths and weaknesses of different techniques, offering valuable information on their performance and applicability in medical imaging. Research contributes to a better understanding of how different generative models perform in the context of pathology.

- **Novel Study of Diffusion Models in Pathology:** The publication not only uses baseline methods to synthesize images but also explores the adaptability of diffusion models by generating images of varying patch sizes. This novel study explores the innovative capabilities of diffusion models, particularly their unique ability to generate images with varying FOVs by prompting them with different patch sizes (4.3). It highlights how these models can produce high-quality synthetic pathology images that accurately reflect different features at different levels. By generating images with varying focal lengths (FOVs), diffusion models can capture

the intricate details and structural variations of tissue samples. This capability can become useful to enhance the versatility and effectiveness of models in research and educational settings.

- **Exploration of Synthetic Dataset Applicability:** This research also explores the applicability of synthetic datasets generated by diffusion models in enhancing deep learning classifiers in Section 4.4. The study evaluates how well these synthetic images can be used to train and improve the performance of deep learning models, specifically in the context of medical image classification. One of the key contributions of this thesis is the demonstration of improved classification performance by incorporating synthetic datasets alongside real images. Empirical evidence is presented, showing that the addition of high-quality synthetic images can significantly boost the performance of classifiers, leading to a more robust classifier.

  This approach not only helps to train more effective deep learning models, but also ensures a more equitable representation of all classes within the dataset, which is crucial for reliable and accurate pathological analysis.

In general, this thesis highlights meaningful contributions to the field of computational pathology by providing a novel and comprehensive study of diffusion generative models, demonstrating their practical applications, and addressing key challenges such as the need for high-quality synthetic datasets.

1.3 Organization of the Thesis

The thesis is organized as follows. Chapter 2 focuses on the initial literature review of diffusion-generative models. In this chapter, we identify various methods that can be applied to pathology datasets. These methods have already been employed and are well known in the Computer Vision domain. After a thorough review, several methods were shortlisted and applied to the computational pathology dataset. Due to the inherent complexity and time sensitivity of diffusion generation models, our approach primarily utilized baseline methods, which offered a simpler and more time-efficient solution. Pathology data itself are complex, involving multiple magnification levels

and fields of view, which will be discussed further in the chapter.

Chapter 2 provides a comprehensive understanding of diffusion generation models, including a review of all the methods examined. Chapters 3 and 4 present the main research contributions, building on the foundational work laid out in Chapter 2.

Chapter 2

Review of Diffusion Generative Methods

This chapter provides an in-depth exploration of the generative techniques exploited in this work starting with their core concepts, advantages, and limitations. The chapter concludes with a review of the related work in the field, highlighting current challenges and potential solutions in the generation of images using these models.

2.1 Introduction

Generative models have become a cornerstone in machine learning, enabling the creation of new data samples that resemble a given dataset. These models have found applications in a variety of domains, including the generation of images (Ho, Saharia, et al., 2022; T.-C. Wang et al., 2018) and videos (Ho, Chan, et al., 2022), the super resolution of images (H. Li et al., 2022), and drug discovery (B. Tang, Ewalt, & Ng, 2021). The importance of generative models lies in their ability to understand and replicate the underlying distribution of data, facilitating advancements in both theoretical and practical aspects of machine learning. Generative models generally must meet several key criteria to be applicable to real-world problems. These criteria include (i) producing high-quality samples, (ii) ensuring mode coverage and sample diversity, and (iii) achieving fast execution times with computationally efficient sampling (Xiao, Kreis, & Vahdat, 2021).

Generative models can be broadly categorized into several types, including Generative Adversarial Networks (GANs), Variational Autoencoders (VAEs), and Diffusion Generative Models. Each

of these models approaches the problem of data generation from different perspectives, offering unique advantages and facing distinct challenges. We will see each of these in detail and how diffusion model can fit in this landscape.

2.2 Diffusion Models in the Generative Model Landscape

Variational Autoencoders (VAEs): (Kingma & Welling, 2013) operate within the realm of likelihood-based models. VAEs leverage variational inference to approximate the target distribution $p(\mathbf{x})$ by minimizing the Kullback-Leibler divergence between the approximate distribution and the target distribution. VAEs consist of an encoder network that maps the input data to a latent space, and a decoder network that reconstructs the data from the latent space. The training process involves optimizing the model to minimize the reconstruction error while ensuring that the latent space follows a specified prior distribution. However, the reliance on simplified prior and posterior distributions, such as Gaussian distributions, may not adequately capture the diversity and nuances present in real-world data, potentially limiting the model's ability to generate high-quality and diverse samples.

Generative Adversarial Networks (GANs): (Goodfellow et al., 2020) Representing a prominent example of implicit generative models, GANs introduce an adversarial framework involving two networks: a generator network that aims to create data indistinguishable from real data, and a discriminator network that strives to differentiate between real and generated data. The training process involves a min-max optimization game, where the generator and discriminator iteratively refine their strategies. Despite their impressive results, GANs are not without challenges. Adversarial training can be unstable and may lead to mode collapse, where the generator produces limited diversity in generated samples (Wiatrak, Albrecht, & Nystrom, 2020).

Both likelihood-based models, such as VAEs, and implicit models, such as GANs, have limitations. VAEs require strong restrictions on model architecture to approximate maximum-likelihood training for tractability, whereas GANs often require adversarial training, which can lead to mode collapse. Diffusion generative models offer a new way to find tractable solutions without adversarial training, falling into the category of likelihood-based generative models.

Diffusion Generative Models (DGMs) represent a relatively new class of generative models that have shown state-of-the-art performance in image generation tasks (Ho et al., 2020; J. Song, Meng, & Ermon, 2020; Y. Song & Ermon, 2020). Unlike GANs and VAEs, DGMs do not require adversarial training, which simplifies the training process and mitigates issues like mode collapse.

DGMs operate by gradually adding Gaussian noise to input data in a process known as the forward diffusion process. This process converts the data into a noise distribution over several time steps. The model then learns to reverse this process, removing the noise step by step to recover the original data distribution. This reverse process is learned using a neural network that predicts the denoised image at each step, ultimately generating high-quality images from random noise.

The absence of adversarial training in DGMs means that they avoid the instability and convergence issues commonly associated with GANs. Additionally, the use of a diffusion process allows DGMs to model complex data distributions more effectively than VAEs, which often rely on simplified assumptions about the data distribution. This makes DGMs particularly suitable for applications that require high-quality, diverse data generation, such as image synthesis and enhancement.

2.3 Theoretical Foundations of Diffusion Generative Methods

This section examines the theoretical underpinnings and structural framework of diffusion generative models (DGMs), specifically denoising diffusion probabilistic models (DDPM) (Ho et al., 2020). We will explore the foundational processes that govern these models, namely the forward diffusion process and the reverse denoising process. In addition, we will examine the mathematical formulations that enable these processes, the training objectives, and the sampling techniques.

2.3.1 Denoising Diffusion Probabilistic Model (DDPM):

Forward Process:

The original image x_0 is slowly corrupted iteratively using a Markov chain by adding scaled Gaussian noise in the forward process.

Given a data point x_0 (original image) sampled from the real data distribution $q(x)$ $x_0 \sim q(x)$, one can define a forward diffusion process by adding a small amount of Gaussian noise to the

sample(x_0) in T steps, producing a sequence of noisy samples $x_1, x_2, \ldots, x_T$. Specifically, at each step of the Markov chain, we add Gaussian noise with variance β_t to x_{t-1}, producing a new image x_t with distribution $q(x_t|x_{t-1})$. This diffusion process can be formulated as follows.

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \mu_t = (\sqrt{1-\beta_t})x_{t-1}, \Sigma_t = \beta_t I) \tag{1}$$

At each timestep t, the parameters that define the distribution of the image $\mathbf{x}_t$ are set as mean $\sqrt{1-\beta_t}\,\mathbf{x}_{t-1}$ and covariance $\beta_t\mathbf{I}$.

Whenever we need a noisy sample at any timestep, we have to perform all the steps until $t-1$ in the Markov chain. To address this, the authors (Ho et al., 2020) corrected the above formula.

Two additional terms are defined as:

$$\alpha_t = 1 - \beta_t \quad \text{and} \quad \bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i \tag{2}$$

$$\begin{aligned}
\mathbf{x}_t &= \sqrt{\alpha_t}\mathbf{x}_{t-1} + \sqrt{1-\alpha_t}\boldsymbol{\epsilon}_{t-1} && \text{;where } \boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \cdots \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \\
&= \sqrt{\alpha_t\alpha_{t-1}}\mathbf{x}_{t-2} + \sqrt{1-\alpha_t\alpha_{t-1}}\bar{\boldsymbol{\epsilon}}_{t-2} && \text{;where } \bar{\boldsymbol{\epsilon}}_{t-2} \text{ is a result of two Gaussians (*).} \\
&= \ldots \\
&= \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon} \\
q(\mathbf{x}_t|\mathbf{x}_0) &= \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1-\bar{\alpha}_t)\mathbf{I})
\end{aligned} \tag{3}$$

where $\epsilon \sim \mathcal{N}(0, I)$. The final expression after t steps is:

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{(1-\bar{\alpha}_t)}\epsilon \tag{4}$$

Here, the term $\boldsymbol{\epsilon}$ is the noise term that is randomly sampled from a standard Gaussian distribution and is first scaled and then added to $\mathbf{x}_{t-1}$.

Reverse Process:

To reverse the process, we sample the conditional distribution $q(\mathbf{x}_{t-1}|\mathbf{x}_t)$. However, estimating this

conditional distribution directly is challenging as it requires the entire dataset. Therefore, we learn a model p_θ to approximate these conditional probabilities to run the reverse diffusion process.

The target distribution is:

$$q(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \tilde{\boldsymbol{\mu}}(\mathbf{x}_t), \tilde{\beta}_t \mathbf{I}) \tag{5}$$

The reverse conditional probability $q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)$ is tractable when conditioned on $\mathbf{x}_0$:

$$q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_{t-1}; \tilde{\boldsymbol{\mu}}(\mathbf{x}_t, \mathbf{x}_0), \tilde{\beta}_t \mathbf{I}) \tag{6}$$

After some calculations, this is the mean and variance we get :

$$\tilde{\beta}_t = \frac{(1-\alpha_t)(1-\bar{\alpha}_{t-1})}{(1-\bar{\alpha}_t)} \tag{7}$$

$$\begin{aligned}
\tilde{\boldsymbol{\mu}}_t &= \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})x_t + \sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)x_0}{1-\bar{\alpha}_t} \\
&= \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\frac{1}{\sqrt{\bar{\alpha}_t}}(\mathbf{x}_t - \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_t); \text{ Rearranging } x_0 \text{ value from equation 4} \\
&= \frac{1}{\sqrt{\alpha_t}}\Big(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_t\Big)
\end{aligned} \tag{8}$$

The learned model p_θ approximates the conditional probabilities for the reverse process:

$$p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)) \tag{9}$$

The parameters of the learned model are optimized to minimize the difference between the predicted noise and the actual noise added in the forward process:

$$L_t^{\text{simple}} = E_{t\sim[1,T],x_0,\epsilon_t}\left[\|\epsilon_t - \epsilon_\theta(x_t, t)\|^2\right] \tag{10}$$

By traversing backward along the learned reverse chain, starting from pure noise, we can generate new samples resembling the original data. This reverse process allows the model to effectively

reconstruct clean data from noisy inputs.

2.4 Categorization of Diffusion Generative Methods

In the realm of diffusion generative models, various techniques have been developed to enhance their performance and efficiency. These methods can be broadly categorized into three groups: Efficent and accelerated sampling and removal of inconsistencies, and different representation spaces.

2.4.1 Efficent and accelerated sampling

Efficient and accelerated sampling techniques aim to reduce the computational burden and time required for generating high-quality images in diffusion models. By optimizing the sampling process, these methods enhance the speed of image generation while maintaining, or even improving, the fidelity and detail of the generated images. The following subsections discuss several techniques designed to achieve these objectives.

Denoising Diffusion Implicit Model (DDIM):

Denoising Diffusion Implicit Models (DDIM) (J. Song et al., 2020), introduced as an extension to Denoising Diffusion Probabilistic Models (DDPM), address one of the main limitations of DDPMs: because of their low sampling rate. DDIMs alter the forward procedure of DDPMs to result in a non-Markovian diffusion process that enables deterministic sampling. In DDPMs, the forward process is a Markov chain that gradually injects Gaussian noise into the data samples at each time step. The typical DDIM adjusts this to depend on x_0 and x_t for x_{t-1}, making the generative process more flexible. This modification allows DDIMs to take a deterministic sample at a later step, accelerating the generation process while preserving sample quality. The core idea of DDIMs is the generalized parameterization of the reverse process. Unlike DDPMs, where the variance schedule is predetermined, DDIMs can modify the variance schedule at the time of inference. This results in the possibility of sacrificing detailed sample heterogeneity in favor of increased speed of further data analysis, and we could notice this in our results as well. It should be noted that

the sampling procedure in DDIMs can be treated as a generalization of DDPMs. When the variance schedule is set to zero, DDIMs sample deterministically and generally do not take as long as DDPMs, which sample stochastically. However, when the variance is set to that of DDPMs, DDIMs can reproduce the original DDPM sampling.

Like DDPMs, DDIMs share the same training objective; thus, preexisting DDPM models can be seamlessly integrated to incorporate DDIM sampling. This backward compatibility is a much appreciated practical benefit, and once those in the field have developed their models, the increase in sampling speed provides an instant advantage.During sampling, only a subset of S diffusion steps $\{\tau_1, \ldots, \tau_S\}$ are sampled during the inference process.

The choice of DDIM variance is given by the parameter η. The diffusion model is a DDIM when $\eta = 0$ as there is no noise and when $\eta = 1$ the diffusion model is DDPM. Any value between 0 and 1 for η is an interpolation between a DDIM and a DDPM.

ERA-Solver : Error Robust Adams Solver for Fast Sampling

The conception of the ERA-Solver, or the Error Robust Adams Solver, is a significant contribution of theoretical thinking to improve the sampling approach in diffusion models (S. Li, Liu, Chai, Li, & Tan, 2023). This method works to solve difficulties with respect to the control of speed and quality in the generation process when the agility of the sampling phase is reduced or minimized. Conventional techniques for selecting sampling points in diffusion models, such as Euler strategies, compromise accuracy when the number of sampling steps is lowered. Such errors may result in a decline in the quality of the input samples and the samples generated in their place. The ERA-Solver addresses this issue by borrowing methods related to numerical analysis, Adams-Bashforth, to diffusion model sampling (S. Li et al., 2023). The general premise of the ERA-Solver is to utilize data from previous steps to estimate the following situation instead of the current one exclusively. This structured approach to building such a model affords more precise forecasts, since the number of steps is reduced or the ratios are higher. (S. Li et al., 2023) state that the ERA-Solver incorporates an error-correcting model with learning capabilities. It determines the local truncation error of the step and can produce a different step size compared to the step size that was used previously. This flexibility allows the solver to step out in areas where the error is small and get more detailed in places

where it is significant. A primary benefit of the ERA-Solver is the feasibility of obtaining accurate solutions earlier with fewer sampling steps than the standard procedure. This can result in orders of magnitude improvements in the speed of the generation process while maintaining the quality of the generated samples. Furthermore, the designed ERA-Solver enables training in existing diffusion models without additional training (S. Li et al., 2023). Thus, it is a versatile solution that can be implemented quickly and helps to enhance the value of numerous diffusion-based generative applications.

Boosting Diffusion Models with an Adaptive Momentum Sampler

In the work of (X. Wang, Dinh, Liu, & Xu, 2023), the authors present a new algorithm for generating the sampling of the diffusion model that relies on optimization tools such as Adam. This method tries to control the fluctuations of the sampling process and the samples produced, especially if several steps are taken during the entire sampling process. The authors notice that the equivalent vanilla reverse process in diffusion models introduces instabilities that show up as high noise in the images or even complete elimination of high-level features in the coupled generated images. They also partly blame this on the Markovian process of the schema, which delivers a novel update direction of a step, but in a blinders' perspective that does not contain a view of the overall update.

To address these issues, the Adaptive Momentum Sampler incorporates momentum and adaptive step sizes. The momentum term helps to smooth out oscillations in the sampling trajectory and the adaptive step size is implemented using a moving average of squared increments, like the RMSprop Hence, with such mechanisms, the Adaptive Momentum Sampler attempts to build high levels of semantic structures, namely shapes and outline accents, as well as low-high detail variants, including refined texture. As stated by (X. Wang et al., 2023), the present technique is more stable, and there is an improvement in sample quality where fewer sample steps are included. As for characteristics, the main advantage of the Adaptive Momentum Sampler is that it does not have to be re-trained and can work with any initial diffusion model that has previously been trained. Thus, it is possible to conclude that it is a valuable and productive tool that increases the efficiency of the different models.

2.4.2 Removing inconsistencies during sampling

This category focuses on addressing and mitigating inconsistencies that may arise during the image generation process. Techniques in this group work to ensure that the generated images are coherent and realistic, reducing artifacts, and improving overall image quality.

Elucidating the Exposure Bias in Diffusion Models

(Ning, Li, Su, Salah, & Ertugrul, 2023) further outline the exposure bias problem in diffusion models and present an approach to solving it. Exposure bias arises because, at training, the model is trained on clean data. At the same time, during the sampling process, it is conditioned on the possibly noisy outputs from the model itself. (Ning, Li, et al., 2023) first capture the position of the sampling distribution when applying the diffusion models. They found that increases in the prediction error for each sampling step lead to exposure bias most severely. The authors also introduce a different method, called 'Epsilon Scaling', which does not require training. This method introduces a scaling factor λ_t to the reverse process:

$$\mu_\theta(x_t, t) = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \beta_t \sqrt{1 - \bar{\alpha}_t} \frac{\epsilon_\theta(x_t, t)}{\lambda_t} \right) \tag{11}$$

The scaling factor λ_t is developed to address the total prediction error in the context of the sample. After that, the authors provide a concise framework for designing the scaling schedule from the L_2-norm quotient of the forecasted noise over the time stamps. Another consideration that arises because of the Epsilon Scaling is that it can be applied at the sampling level, and whatever change is being made is not influenced during the training phase. As such, it qualifies as a simple and flexible way of improving the quality of samples from diffusion models with pre-trained data. (Ning, Li, et al., 2023) reveal that exposure bias can be reduced successfully using Epsilon Scaling, and this helps enhance the quality of the sample, mainly when a few steps are being used to sample. This makes it especially useful in samples where the rate is faster.

Input Perturbation Reduces Exposure Bias in Diffusion Models

The Input Perturbation method introduced by (Ning, Sangineto, Porrello, Calderara, & Cucchiara, 2023) is designed to solve one of the significant problems of diffusion models – exposure bias. (Ning, Sangineto, et al., 2023) describe exposure bias as when the model is trained with clean data. Still, during sampling, its prediction passes through a non-deterministic post-processing step, exposing the model to its own possible mistakes. These differences can result in the deterioration of the quality of the samples collected, mainly when there are many alternative sampling steps. The principle of the Input Perturbation method is to replicate all the errors that may occur when using the sampling method while training the model. This is done by including noise, which is controlled during the training of the inputs. Specifically, the method modifies the forward process of diffusion models as follows:

$$y_t = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t}(\epsilon + \gamma_t \xi)$$

Here, ξ is an additional noise and γ_t is the scaling factor for this additional extra noise at time step t. Instead of the regular noisy sample, x_t, the perturbed input y_t is fed into the training process as shown below. This is because, through the process explained in the Input Perturbation method, the original inputs are trained to be less sensitive to the types of errors that are likely to occur during the sampling process (Ning, Sangineto, et al., 2023). However, this increased robustness sometimes results in a better sample most of the time in conditions that entail fewer steps in creating the sample.

Therefore, the most significant advantage of this method is that one does not need time or any particular setup for preparation. Thus, it can be readily incorporated into the model architectures and the training regimes derived from the traditional diffusion models. Besides, the required modification of the existing classifier is minimalist; there is an additional scalar hyperparameter, which is the variance of the extra noise.

2.4.3 Different representation spaces

Diffusion models traditionally operate in the pixel space, which is computationally intensive and may not always yield the most efficient results. To address these challenges, alternative representation spaces, such as latent spaces and gradient domains, have been explored. These methods

aim to improve the efficiency and effectiveness of diffusion models by reducing the dimensionality of the data they operate on, leading to faster processing times and potentially higher quality outputs. By leveraging these different representation spaces, models can achieve significant computational savings and enhance their ability to generate detailed and realistic images.

Latent Diffusion Model(LDM)

In their study, (Rombach et al., 2022) posited that LDMs are an improvement over the diffusion models and embrace efficiency and scalability. LDMs' primary advancement is that the diffusion process is quantified by the learned latent distribution instead of pixel distribution. The LDM architecture consists of three main components: an encoder, decoder, and U-Net, developed and arranged as the diffusion model. Such transformations enable the encoder to map the input images to a lower-dimensional latent space, while the decoder's task is to map the obtained representations back to the image space. It should be noted that the diffusion process occurs solely within this learned latent space. The training process of the LDMs consists of two unique phases. First, an autoencoder is trained such that this model can generate a small feature vector that can represent the input data (Rombach et al., 2022). Then, a diffusion model is trained in this latent space, meaning it needs to reverse the process from the data space to the latent space. The latent diffusion process's training objective is similar to pixel-space diffusion models. Still, it operates on the latent representations:

$$LDM := E_{E(x),\epsilon\sim\mathcal{N}(0,1),t}\|\epsilon - \epsilon_\theta(z_t, t)\|_2^2. \tag{12}$$

E is the encoder function, and z_t is the noisy latent representation for a specific time step, t. Using LDMs is an advantage since they require much less timesteps than pixel-space diffusion models due to the optimization process, which means it converges faster in latent space. Given that LDMs work in a lesser dimension latent space, producing high-quality images becomes significantly more efficient. This makes it possible to generate diffusion models on extensive data and high resolutions compared to earlier models. LDMs have proven effective across multiple synthesis-based operations, including conditional generation and image-to-image translation.

Gradient Domain Diffusion Models for Image-Synthesis

According to (Gong, 2023), Gradient Domain Diffusion Models develop a new idea of applying the diffusion process to the gradient domain rather than the pixel space. Using the gradients of images, this method seeks to potentially enhance the quality and sampling rate of the model through the sparse and structured format inherent in this element. The primary concept of this approach lies in the modification of the forward and reverse processes of the diffusion models in terms of image gradients (Gong, 2023). The main advantage of the gradient domain is that an analyst will see that the image gradients are comparatively sparse. Therefore, it is much easier for the learning process in the model to manage them. In addition, the gradient domain is more comprehensive and includes edge and texture that significantly impact image quality, as highlighted in (Gong, 2023). When generating images using the learned gradient model, sums of these gradients are required to get pixel values for the image. This is usually accomplished using some Poisson solver or a learned integration network (Gong, 2023). The real-valued models, specifically the Gradient Domain Diffusion Models, have been observed to perform better in image synthesis than some pixel-space models with better image quality and definition. Nevertheless, they also bring more complexity through the availability of the gradient's computation and integration parts.

2.5 Discussion and Concluding Remarks

In conclusion, Chapter 2 has laid a comprehensive foundation by examining the landscape of Diffusion Generative Models and their applications across various domains. This literature review serves as a crucial basis for the next phase of our research, where we will identify and implement the most promising baseline methodologies on pathology datasets, our primary area of interest. By thoroughly exploring and evaluating the available options, we are well-prepared to embark on the detailed experimental work and innovation presented in the subsequent chapter.

Chapter 3 and 4 will build upon the foundational work laid out in Chapter 2 by introducing the dataset and pre-processing steps necessary for our research. Chapter 3 will detail the KGH dataset, annotation processes, extraction of Regions of Interest (ROIs), and the methodologies applied to prepare the data for analysis. Chapter 4 sets the stage evaluation of Diffusion Generative Models on

this dataset, leading to the significant advancements in computational pathology.

Chapter 3

Dataset and Methodology

This chapter provides a detailed overview of the dataset used in this study, specifically the Kingston General Hospital (KGH) dataset, and outlines the pre-processing steps undertaken to prepare the data for analysis using Diffusion Generative Models (DGMs).

We begin by introducing the KGH dataset, highlighting its relevance and unique features that make it suitable for this research. The process of annotation and extraction of Regions of Interest (ROIs) is then discussed, detailing how critical areas within the images were identified and marked. This is followed by an explanation of the FOV and resolution levels, emphasizing the importance of capturing relevant image regions at appropriate scales and the resolution adjustments made to manage the high-dimensional nature of pathology images.

In addition, we explore the specifics of the annotations and cancer tissue classes, providing information on how different tissue types were categorized and labeled. The chapter concludes with a description of the methodology applied to preprocess the dataset, setting the stage for the implementation of DGMs in subsequent chapters.

3.1 KGH Dataset

The Kingston General Hospital (KGH) dataset comprises **1037** whole slide images (WSI) of healthy colon tissue and colon adenocarcinoma. Colorectal adenocarcinoma, the most common type of colorectal cancer, accounts for 95% of all cases. It originates in the cells of the intestinal

glands that line the colon and rectum, typically beginning as small polyps that can develop into cancer if not treated. The KGH dataset includes healthy colon tissue and four types of colon polyps: Tubular adenoma (TA), sessile serrated lesions (SSL), hyperplastic polyps (HP) and tubulovillous adenoma (TVA).

For these four tissues, the pathological slides in the KGH dataset are annotated at the Region of Interest (ROI) level. Annotations for the dataset were provided by Dr. Sonal Verma. It is important to note that not all regions outside the annotated areas are normal; they are simply not annotated. For our study, we focus exclusively on annotated regions of cancer tissue classes such as HP, SSL, TA, and TVA.

Figure 3.2 illustrates how ROIs are identified and highlighted in blue within the entire slideshow. These WSIs are rectangular, with sizes around 60,000 $\times$ 100,000 pixels, requiring approximately 1 GB of memory each.

The KGH dataset includes healthy colon tissue and four types of colon polyps:

- **HP (Hyperplastic Polyps):** Non-cancerous growths with a small risk of malignant transformation (Jass & Burt, 2000), characterized by cellular overgrowth. These polyps are often found in the distal colon and are generally less than 5 mm in diameter. Despite their benign nature, regular monitoring is recommended to ensure that they do not develop cancer.

- **SSL (Sessile Serrated Lesions):** Pre-cancerous lesions that are flat or slightly elevated (Murakami et al., 2022), making them difficult to distinguish from HP (Hyperplastic Polyps) visually. They have a distinctive histological appearance with a serrated architecture and are considered high-risk due to their potential to progress to colorectal cancer if left untreated.

- **TA (Tubular Adenomas):** Small, benign polyps with a 10% chance of transforming into cancer. These adenomas (Salemis et al., 2012) are the most common type of polyp found during colonoscopies. They usually appear as small, pedunculated (stalked) growths and are composed of tubular glands. TAs (Salemis et al., 2012) are generally considered low-risk, but larger adenomas or those with high-grade dysplasia require closer surveillance (Jaravaza & Rigby, 2020).

- **TVA (Tubulovillous Adenomas):** Lesions with the highest potential to transform into cancer.

TVAs have a mixed histological structure, containing both tubular and villous elements. They are typically larger than other adenomas and have a higher risk of malignancy.

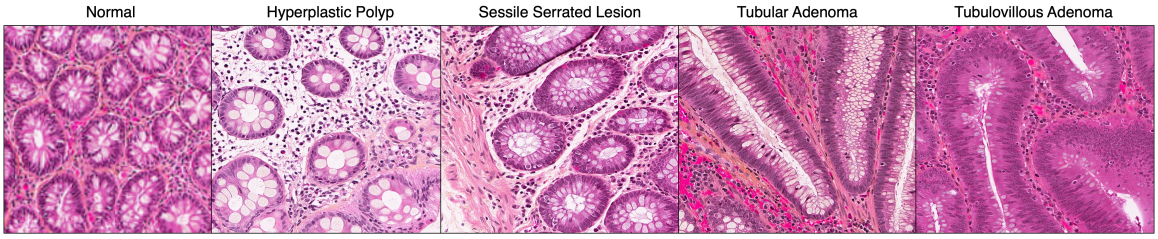


Figure 3.1: Samples from Cancer Tissue and Normal WSI

We can visualize patches from these classes in Figure 3.1. Despite some visual differentiation in certain regions amond these classes, distinguishing them can be challenging without a pathologist's expertise. Understanding these classes is crucial for making accurate analogies with our generated images. The visual distinctions, although subtle, provide necessary context for training our models.

3.1.1 FOV and Resolution level

The KGH dataset contains images at four resolution levels: 20x, 5x, 1.25x, and 0.3125x. These levels provide varying perspectives of the same tissue, which is crucial for comprehensive pathological analysis. The high resolution of these images allows for detailed examination of cellular structures, which is essential for accurate diagnosis and research.

To manage the high-dimensional nature of these pathology images, we extract smaller sections referred to as patches. To extract these patches, we specify the FOV. We also eliminated noisy images from the original set.

3.2 PKGH Creation

3.2.1 Input FOV and Patch Size

To extract patches from the WSIs, we specified the FOV and patch size based on the resolution levels. The primary FOVs used were 224 and 336, resized to 128 $\times$ 128 and 192 $\times$ 192 pixels respectively. This selection was made to balance the need for sufficient cellular structure and the availability of a sufficient number of patches for robust training.

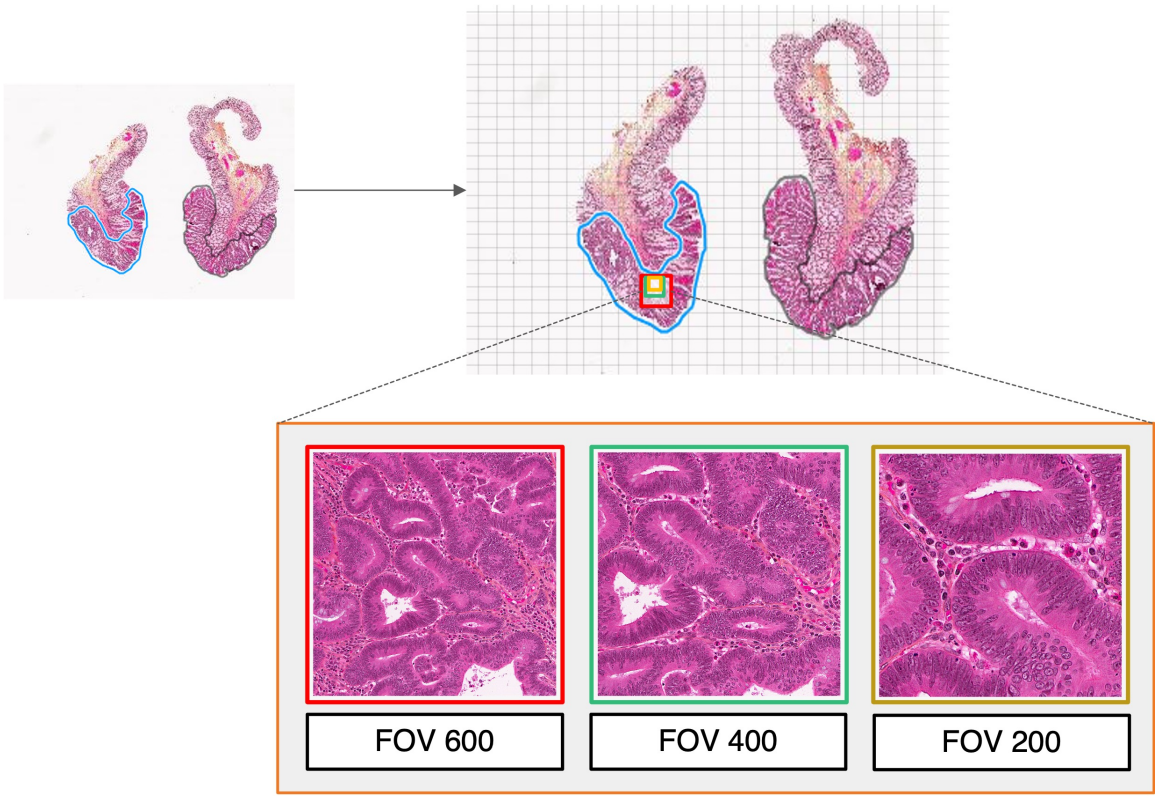


Figure 3.2: The images demonstrate how Regions of Interest (ROI) are annotated and highlighted (shown in blue) on the left side, followed by the extraction of different patches captured at FOV 600, FOV 400, and FOV 200 on the right side.

3.2.2 Pipeline used to create PKGH

The creation of the PKGH dataset involved the following steps:

(1) **Annotation Extraction:** We utilized a code developed by my lab mate, Cassandre Notton, to extract regions of interest (ROIs) from WSI using the TIAToolbox (Pocock et al., 2022) to extract ROIs from colon polyps: Hyperplastic Polyps (HP), Sessile Serrated Lesions (SSL), Tubular Adenomas (TIA), and Tubulovillous Adenomas (TVA). Healthy colons are just simply used directly for extraction without annotations.

(2) **Patch Extraction:** Patches were extracted from the annotated WSIs at specified Fields of View (FOVs). For our experiments, we maintained a consistent patch resolution of 1.75. This approach allowed us to observe how different FOVs provide various perspectives of the same tissue at the same magnification. Specifically, we used FOVs of 224 and 336 for creating the

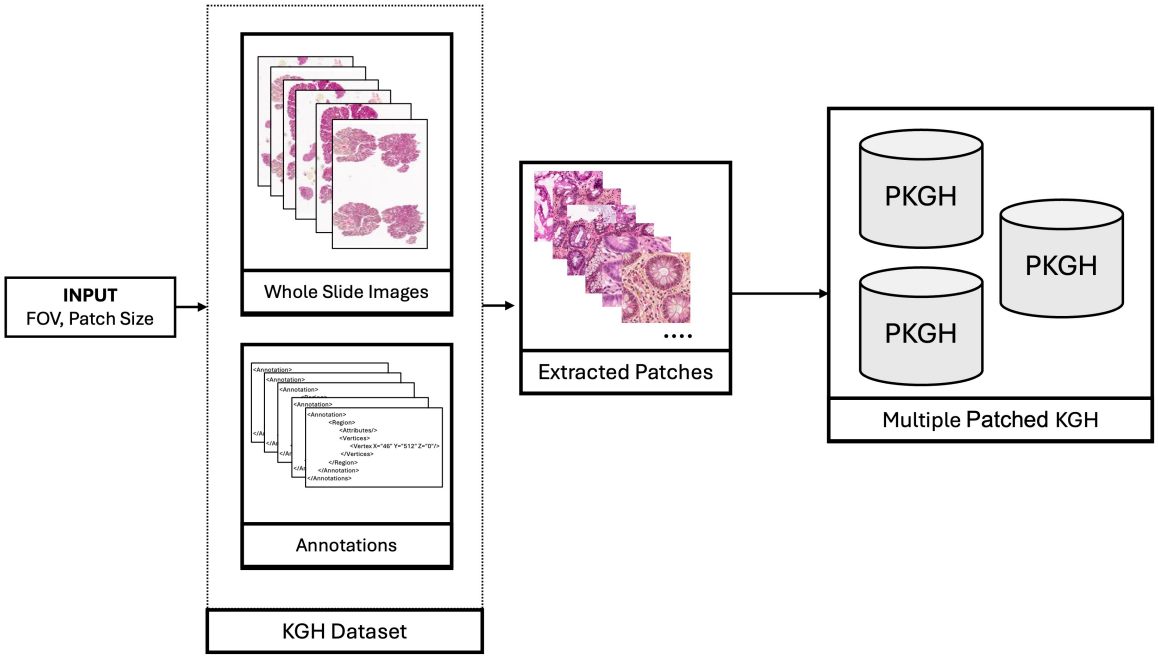


Figure 3.3: PKGH Creation from original KGH Dataset

PKGH dataset. These FOVs were selected to ensure that the patches offered a comprehensive view of the cellular and tissue structures necessary for training generative models. The chosen FOVs are not excessively large to avoid having a limited number of patches available for training and classification.

From the training data, we obtained approximately 50,000 images (about 10k in each class). We used 80% of the WSIs for training diffusion models, ensuring a robust dataset for model development. The remaining 20% of the WSIs were reserved for testing. From these test WSIs, we extracted 10,000 images for testing, ensuring they are never-seen images for the diffusion models. This split helps in validating the performance and generalizability of the deep learning classifier models.

This pre-processed dataset ensures that we have a comprehensive and well-annotated collection of images, enabling us to train our diffusion generative models effectively.

3.3 Methodology

Diffusion models (Ho et al., 2020; Sohl-Dickstein, Weiss, Maheswaranathan, & Ganguli, 2015; Y. Song & Ermon, 2019) are a powerful class of probabilistic generative models that are used to learn complex data distributions. The purpose of this framework is also to demonstrate how various methods can be integrated seamlessly. We explain each category and method in a manner that allows any standalone model or technique to be easily plugged into the framework. The framework is explained as follows:

(1) **Training:**

- **Forward Process:** Gaussian noise is introduced into the input images, with the magnitude determined randomly by a scheduler, resulting in varying levels of noise. This allows the model to explore a range of image noises, from subtle distortions to isotropic Gaussian noise. The noisy image, along with embedding vectors representing the noise level and class labels, is fed into a U-Net-based neural network.
- **Reverse Process:** The U-Net predicts the noise added to the images. This prediction is compared to the actual noise to calculate the loss, guiding the model's training process.

(2) **Sampling:**

- **Reverse Process:** The process starts with pure isotropic noise. The trained network predicts the corresponding noise level, and the difference between the predicted noise and the initial noise creates a less noisy image. This image is iteratively refined through multiple steps to generate the final image.

This whole process is explained well in the figure 3.4.

3.3.1 Denoising Diffusion Probabilistic Models

Denoising Diffusion Probabilistic Models (DDPM) (Ho et al., 2020) follow the same training and sampling methods explained above, with both the forward and reverse processes occurring in

Table 3.1: Overview of Diffusion Model Methods and Their Integration

Method	Can be used as a plugged-in method?	Usage in Framework	Pixel or Latent space	Special Features
DDPM 3.3.1	Yes	Training and Sampling	Pixel space	Standard diffusion model
LDM 3.3.2	No	Training and Sampling	Latent space	Uses autoencoder for latent space, cross-attention
CFG 3.3.3	Yes	Training and Sampling	Depends on baseline model	Allow conditional sampling without separate classifier
DDIM Sampling 3.3.4	Yes	Sampling method	Depending on baseline model	Non-Markovian process, subset sampling
Epsilon Scaling Sampling 3.3.4	Yes	Sampling method	Depending on baseline model	Scales epsilon value to reduce exposure bias

the pixel space. In the forward process, Gaussian noise is added to the image x_0 at each timestep t according to:

$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I}) \tag{13}$$

where β_t is the variance schedule. The reverse process aims to denoise the image step-by-step using a learned model p_θ, represented as:

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t)) \tag{14}$$

The objective function to train the model is:

$$L_t^{\text{simple}} = E_{t\sim[1,T],\mathbf{x},\boldsymbol{\epsilon}_t}\left[\|\boldsymbol{\epsilon}_t - \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)\|_2^2\right] \tag{15}$$

3.3.2 Latent Diffusion Models

Latent Diffusion Models (LDM) (Rombach et al., 2022) operate in a lower-dimensional latent space, which is learned by an autoencoder. The training and sampling stages follow the same principles as the general diffusion models described above, but with processes occurring in the latent space. An extra layer is added to encode images into this latent space before the diffusion processes begin. The forward process is defined as:

$$q(\mathbf{z}_t|\mathbf{z}_{t-1}) = \mathcal{N}(\mathbf{z}_t; \sqrt{1-\beta_t}\mathbf{z}_{t-1}, \beta_t\mathbf{I}) \tag{16}$$

where $\mathbf{z}_t$ represents the latent variables. The reverse process denoises the latent variables step-by-step:

$$p_\theta(\mathbf{z}_{t-1}|\mathbf{z}_t) = \mathcal{N}(\mathbf{z}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{z}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{z}_t, t)) \tag{17}$$

LDMs also incorporate cross-attention mechanisms within the architecture, enhancing conditional image synthesis. After the latent space processing, the autoencoder transforms the latent variables back into images. The objective function for training remains similar:

$$L_t^{\text{simple}} = E_{t\sim[1,T],\mathbf{z},\boldsymbol{\epsilon}_t}\left[\|\boldsymbol{\epsilon}_t - \boldsymbol{\epsilon}_\theta(\mathbf{z}_t, t)\|_2^2\right] \tag{18}$$

3.3.3 Conditioning

Recent advancements in DPMs have introduced class-conditional generation, where additional class-related information is incorporated to guide the generation process. Our findings revealed that class-conditional generation significantly enhances the fidelity of the generated images to specific class characteristics. As the guidance scale increased, the generated images exhibited more precise and accurate representations of the target classes.

Classifier-Free Guidance

Classifier-Free Guidance (CFG) (Ho & Salimans, 2022) is a technique that enables the generation of high-quality samples without relying on a classifier, addressing the limitations associated with classifier guidance. CFG modifies the score function in a way that emulates the effects of classifier guidance, but without using an explicit classifier. The approach involves training an unconditional denoising diffusion model alongside the conditional model, using a single neural network to parameterize both. The sampling process utilizes a combination of the conditional and unconditional score estimates, allowing for effective guidance without a classifier. This results in the production of high-quality synthetic images that are both varied and representative of the original dataset, enhancing the model's performance in generating realistic images.

$$\epsilon_t = (1 + w) * \epsilon_\theta(x_t, c) - w * \epsilon_\theta(x_t) \tag{19}$$

Here, $\epsilon_\theta(x_t, c)$ is conditional model and $\epsilon_\theta(x_t)$ is unconditional model. w is used as a guidance scale.

3.3.4 More Sampling Choices

To explore different sampling methods and their effects on image generation and model performance, we incorporated two additional techniques: Denoising Diffusion Implicit Model (DDIM) and Epsilon Scaling. DDIM accelerates image generation by introducing a non-Markovian process that redefines the diffusion process, utilizing a subset sampling strategy for faster sampling without compromising model performance. The reverse diffusion process in DDIM is defined as:

$$\mathbf{x}_{t-1} = \sqrt{\alpha_{t-1}} \left(\frac{\mathbf{x}_t - \sqrt{1-\alpha_t}\epsilon_\theta(\mathbf{x}_t)}{\sqrt{\alpha_t}} \right) + \sqrt{1-\alpha_{t-1}-\sigma_t^2}\epsilon_\theta(\mathbf{x}_t) + \sigma_t \epsilon_t \tag{20}$$

where the variance σ_t is given by:

$$\sigma_t = \eta \sqrt{\frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t} \cdot \beta_t} = \eta \sqrt{\tilde{\beta}_t} \tag{21}$$

In DDIM, setting $\eta = 0$ eliminates noise, making it equivalent to DDPM, while $\eta = 1$ maintains the standard diffusion model, allowing interpolation between DDIM and DDPM. To address exposure bias, we employed Epsilon Scaling (Ning, Li, et al., 2023), which scales the epsilon value using a linear function, ensuring consistency between training and sampling, thereby reducing bias and improving sample quality. The epsilon value is scaled as:

$$\epsilon_t = \frac{\epsilon_\theta(\mathbf{x}_t, t)}{\lambda_t} \tag{22}$$

where $\lambda_t = \lambda_t k + b$. The scaled ϵ_θ is then used in the DDPM equation:

$$\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} \boldsymbol{\epsilon}_t \right) + \sigma_t z \tag{23}$$

3.4 DGM Pipeline for Computational Pathology

3.4.1 Training Model and Generative Approach

The training and generative approach for pathology data employs Diffusion Generative Models, as illustrated in Figure 3.4. This framework includes both the training and sampling phases essential for generating high-quality synthetic pathology images. We utilized various diffusion models and sampling methods to compare and study their effects on the generated images' quality and efficiency.

During the **Training phase** (top-right of the figure), the data points progress from x_0 to x_T with noise levels ϵ_0 to ϵ_T. Additionally, we employ class information from different tissue types (Normal, HP, SSL, TA, TVA) to guide the training process. This class information helps the model learn and understand the distinct characteristics of each tissue type.

At **sampling time**, we prompt the model with a class label to generate images corresponding to that specific tissue type. This allows the model to produce synthetic images that accurately reflect the features of the given class.

3.4.2 The Architecture

We used a UNet model based on (Dhariwal & Nichol, 2021) for our experiments. We converted 128 $\times$ 128 images to noise, which were then fed into the UNet network. The network predicted the amount of noise added to these images and outputted the noise prediction. We employed a loss function that compares the predicted noise with the actual noise from the stochastic process. In Diffusion Models, Gaussian noise is iteratively added to the original image according to a variance schedule with a large total number of steps (T = 1000). The UNet model also incorporates timestep and class embeddings to learn class-specific features.

We implemented the LDM architecture from Rombach et al. (2022), which consists of the Variational Autoencoder (VAE), the U-Net denoiser, and added the class embedding. We used an autoencoder based on vector quantization, VQ-autoencoder, referred to as VQF4-DM (Van Den Oord, Vinyals, et al., 2017).

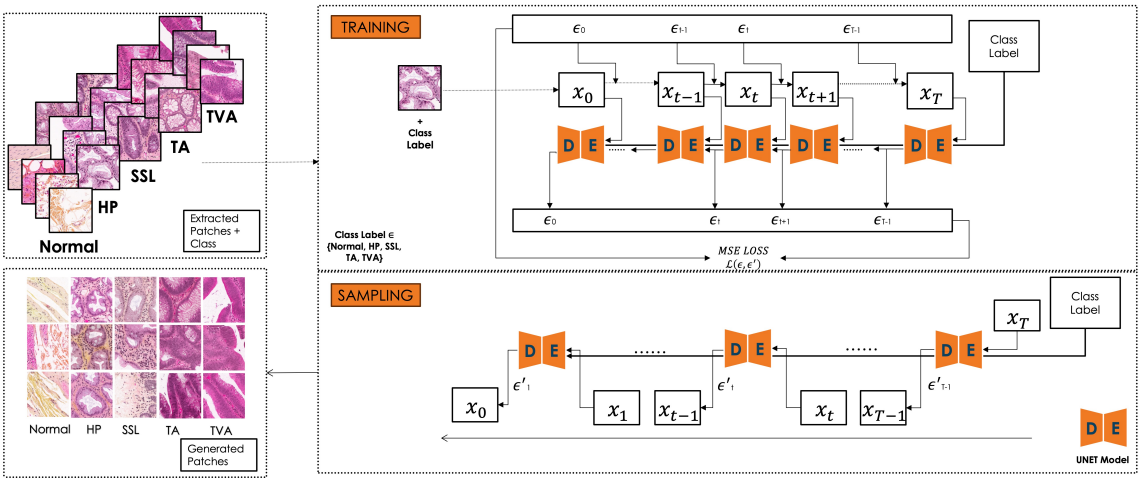


Figure 3.4: Diffusion Generative Models framework used in the pathology dataset: The Training phase (top-right) details the data points' progression from x_0 to x_T with noise levels ϵ_0 to ϵ_T, where the model adds noise to the original data and estimates the reverse process guided by class information to understand tissue details. The Sampling phase (bottom-right) reverses the diffusion from a noisy state x_T to the original data point x_0, by iteratively denoising using noise predictions ϵ' and learned parameters, and prompting the class label to generate images of specific tissue types.

3.5 Evaluation Metrics

A classic approach for generative model's evaluations is to compare the log-likelihoods of models. This approach, however, has several shortcomings. A model can achieve high likelihood, but low image quality, and conversely, low likelihood and high image quality. The two most common GAN (Goodfellow et al., 2020) evaluation measures are Inception Score (IS) (Salimans et al., 2016) and Fréchet Inception Distance (FID) (Heusel, Ramsauer, Unterthiner, Nessler, & Hochreiter, 2017).

Inception Score (IS) rely on a pre-existing classifier (InceptionNet) (Szegedy et al., 2015) trained on ImageNet. IS computes the KL divergence between the conditional class distribution and the marginal class distribution over the generated data. IS does not capture intra-class diversity, is insensitive to the prior distribution over labels (hence is biased towards ImageNet dataset and Inception model. Therfore, we would not be using it for our dataset as it has completely different data than Imagenet. However, we would be using FID and KID.

FID (Heusel et al., 2017) calculates the Wasserstein-2 (a.k.a Fréchet) distance between multivariate Gaussians fitted to the embedding space of the Inception-v3 network of generated and real

images. The Kernel Inception Distance (KID) (Bińkowski, Sutherland, Arbel, & Gretton, 2018) aims to improve on FID by relaxing the Gaussian assumption. KID measures the squared Maximum Mean Discrepancy (MMD) between the Inception representations of the real and generated samples using a polynomial kernel. This is a non-parametric test so it does not have the strict Gaussian assumption, only assuming that the kernel is a good similarity measure.

We further use deep learning based classifiers approach to compare original vs generated data to show how synthetic data is actually applicable in real life. This is separately explained in Section.

Chapter 4

Diffusion Models in Computational Pathology : Image Analysis and Comparative Study

This chapter presents the core research contributions of this thesis, focusing on the application of diffusion-generative models (DGMs) to computational pathology. The primary objective is to analyze and compare the performance of different DGM methodologies in pathological data sets in terms of image synthesis and enhancement.

We begin by describing the experimental setup. Following this, we analyze the results, highlighting key findings and observations. The results section compares the effectiveness of various DGM techniques in generating high-quality pathological images, exploring their strengths and limitations.

The chapter concludes with an evaluation of the pathological images generated, assessing their quality and accuracy. This evaluation validates the performance of the DGM and ensures that the generated images meet the required standards for practical application in medical research and diagnostics. At the end of this chapter, we discuss these findings and significant advances in the use of DGMs for computational pathology.

4.1 Experiment Details

4.1.1 Comparative analysis

In our research, our aim is to comprehensively analyze the effects of various diffusion generation models (DGM) for the pathology data set. Rather than solely identifying the best method, our objective is to facilitate meaningful comparisons between different methods, emphasizing their visual impact and utility, particularly when dealing with large-sized images.

Our comparative study includes two primary training models: Denoising Diffusion Probabilistic Models (DDPM) and Latent Diffusion Models (LDM). These models serve different purposes, and we are interested in exploring how each performs under pathological data conditions and what unique contributions they can make in data synthesis. Latent Diffusion Models (LDMs) offer efficient image generation from the latent space with a single network pass. LDMs operate in the learned latent space, which exhibits better scaling properties with respect to spatial dimensionality. Therefore, latent models are more efficient compared to pixel-based designs.

Notably, the difference between LDM and DDPM lies in LDM's ability to model long-range dependencies within the data by constructing a low-dimensional latent representation and diffusing it. In contrast, DDPMs apply the diffusion process directly to the input images. This capability is particularly useful for medical image augmentation tasks that require capturing complex patterns and structures.

In our implementation, denoising diffusion probability models (DDPM) and latent diffusion models (LDM) differ in their architectures and efficiency in generating high-quality images. DDPMs create detailed and realistic images by iteratively denoising through 500 timesteps in the pixel spaces. Conversely, LDMs, although large in model size, leverage a trained latent space, allowing them to generate images in significantly fewer timesteps, only 200 steps, thus speeding up the generation process while maintaining high image quality. This efficiency makes LDMs advantageous for applications that require rapid image synthesis. We will see the effect of this in the results section 4.2.1

Classifier guidance is a recently developed strategy for balancing mode coverage and sample fidelity in post-training conditional diffusion models. It works similarly to low-temperature sampling

or truncation in other forms of generative models. For example, in recent research, the classifier has undergone training using images corrupted by noise. Subsequently, during the diffusion sampling process, the classifier gradients are utilized to guide the sample toward the intended classification label. Classifier guidance provides a trade-off: It enhances compliance with the conditioning signal and improves overall sample quality, but it can also incur high computational costs for high-quality samples.

We train both DDPM and LDM models on the same pathology datasets and use classifier-guidance strategy. This allows us to observe and evaluate the behavior of these models through visual assessments of the sampled images. Our meticulous approach ensures that we can identify subtle changes and improvements, which is crucial for analyzing large pathological images.

4.1.2 Patch Size

In addition, our experiments introduced a novel hyperparameter: **Patch Size**. By varying the patch size exclusively during the image generation phase, we observed notable differences in tissue structure. The results demonstrated that adjusting the patch size can significantly enhance the model's ability to capture fine details and intricate structures within the generated images according to the patch size we provide. This process is not just simple up-sampling or down-sampling of the patch; rather, the model is able to generate details according to the patch size, effectively learning the patch resolution, further contributing to the robustness and accuracy of our **Histology Patch Generation Model**.

To understand patch resolution, we can look at the histopathology data. Patch resolution in histopathology images refers to the level of detail captured in each patch relative to the actual size of the tissue being imaged. It is typically measured in microns per pixel (mpp). This means that the resolution indicates how many microns of actual tissue are represented by each pixel in the image. For example, a patch resolution of 1.75 mpp means that each pixel in the patch corresponds to 1.75 microns of the actual tissue.

High patch resolution (lower mpp values) captures more detailed information, such as cellular and subcellular structures, making it suitable for identifying fine morphological features. Conversely, lower patch resolution (higher mpp values) provides a broader view of the tissue, capturing

larger structures and the overall tissue architecture. Understanding and adjusting patch resolution is crucial to optimizing histopathological image analysis, as it allows for a balance between detail and context, allowing better identification and classification of various tissue types and pathological changes.

$$\text{FOV} = \text{Patch Size} \times \text{Patch Resolution}$$

We validated our findings by first confirming the patch resolution used earlier, where the FOV was 224 and the patch size was 128, resulting in a patch resolution of 1.75 microns per pixel (mpp) according to the formula mentioned above. Using this patch resolution (1.75 mpp) and varying the patch sizes, we calculated the corresponding FOVs:

Patch Size	**FOV**
64	112
96	168
128	224 (original)
160	280
192	336
224	392

Table 4.1: Patch Sizes and Corresponding Fields of View

We extracted these patches from the original dataset used for WSIs diffusion training. These patches were taken from the same WSIs.

4.1.3 Evaluation of Synthetic Pathology Dataset

The evaluation of generated pathology images is crucial to determine their quality and similarity to real pathology images. Although evaluation metrics provide valuable information about the quality of generated images, their applicability to real datasets remains uncertain. This study aims to analyze the performance of the generated data set on the ResNet-50 architecture (He, Zhang, Ren, & Sun, 2016) along with the real data set.

4.2 Results and Comparative Analysis

The FID scores of our DDPM method demonstrate strong performance when compared with similar studies in the field. These comparisons are crucial for highlighting the effectiveness of our approach in generating high-quality images. By evaluating FID scores across different studies, we can understand the relative success of our method in capturing the complexity and diversity of tissue samples.

The following table compares our FID scores with those of other relevant studies that use the same DDPM methodology. This allows for a direct comparison of performance metrics and provides a clear picture of where our method stands in the context of current research.

The Table 4.2 compares the FID scores of our DDPM method with several studies in the literature, as referenced by (Pozzi et al., 2023). We only consider these among others as these are based on the same DDPM methodology and architecture:

Study	**Number of Classes**	**FID**
Current Study (PKGH_224)	5	**19.08**
Current Study (PKGH_336)	5	**18.45**
(Pozzi et al., 2023)	5	35.11
(Moghadam et al., 2023)	1	20.11

Table 4.2: Comparison of FID scores with existing studies

When compared to other studies, our DDPM method's FID scores of 19.08 for the PKGH_224 dataset and 18.45 for the PKGH_336 dataset are competitive, demonstrating superior performance over (Pozzi et al., 2023) and comparable results with (Moghadam et al., 2023). These results highlight the importance of conditioning in DDPM and the significant impact of the FOV used. The findings underscore the value of FOV selection to achieve better outcomes in histopathological image generation.

4.2.1 Comparative Analysis

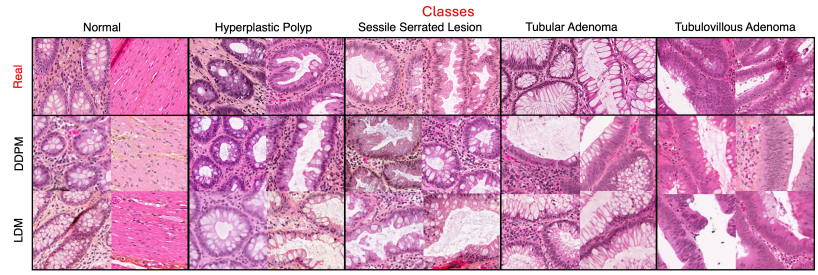


Figure 4.1: Real vs Generated Images from the FOV 224 (DDPM and LDM): The top row shows a representative image from the real dataset for each tissue type, while the bottom row displays a generated image of the same tissue.

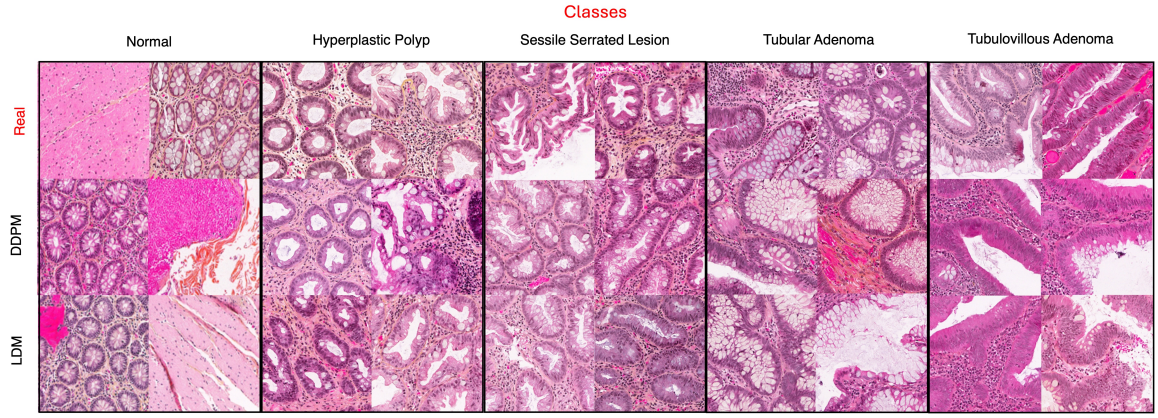


Figure 4.2: Real vs Generated Images from the FOV 336 (DDPM and LDM): The top row shows a representative image from the real dataset for each tissue type, while the bottom row displays a generated image of the same tissue.

The analysis focuses on comparing the performance of Denoising Diffusion Probabilistic Models (DDPM) and Latent Diffusion Models (LDM) across two datasets: PKGH_224 and PKGH_336.

The Table 4.3 below summarizes the results of all methods.

For the PKGH_224 dataset, the DDPM model with DDPM sampling yielded an FID score of 19.08 and a KID score of 0.0134. This indicates a reasonable level of image quality and diversity generated by the model. In comparison, LDM with DDPM sampling yielded an FID score of 24.43 and a KID score of 0.0185. The latent space representation in LDM provides good results, although the DDPM model performed slightly better in this instance. The DDIM sampling method for DDPM had an FID score of 22.66 and a KID score of 0.0154, whereas LDM with DDIM sampling achieved

an FID score of 25.56 and a KID score of 0.0161. Here, DDPM outperformed LDM, highlighting the variations in performance across different sampling methods.

For the PKGH_336 dataset, DDPM with DDPM sampling produced an FID score of 18.45 and a KID score of 0.0129, showing consistent performance across different dataset sizes. LDM with DDPM sampling yielded an FID score of 23.10 and a KID score of 0.0160. The latent space approach in LDM continues to produce high-quality images but did not outperform DDPM in this dataset. The DDIM sampling method for DDPM had an FID score of 21.34 and a KID score of 0.0159, while LDM with DDIM sampling achieved an FID score of 26.41 and a KID score of 0.0199. Again, DDPM demonstrated better performance compared to LDM.

This shows that for the smaller dataset, PKGH_224, DDPM showed a slight edge in image quality and diversity over LDM, particularly with DDPM sampling, which yielded better scores. This suggests that DDPM is more effective in generating detailed and realistic images for this dataset size. In contrast, for the larger dataset, PKGH_336, DDPM again outperformed LDM, consistently producing higher quality images. Both DDPM and LDM demonstrated high performance, but DDPM maintained better scores across different sampling methods, indicating its robustness in handling varying dataset sizes. Overall, the study highlights DDPM's superior performance in generating realistic and diverse synthetic pathology images.

Notably, LDM usually excels in complex conditioning scenarios, such as text and images, making it highly suitable for multimodal generative tasks. However, LDM performs exceptionally well with class conditioning, producing high-quality results despite its optimization for complex inputs. In summary, while DDPM provides a solid foundation, LDM offers faster convergence and comparative FID scores, making it a powerful tool in generative modeling. However, DDPM has been proven to be a useful tool to control pixel space and would also lead us to experiment with different patch size generation in Section 4.3.

Figure 4.3 shows a visual comparison of the results of the DDPM and Epsilon sampling methods applied to histopathology images. The upper row illustrates output from DDPM, while the lower row showcases Epsilon sampling results. Both methods can capture larger tissue structure from pathology images, but the DDPM images display more consistent structural details across different samples. In contrast, Epsilon sampling produces slightly varied textures, which could indicate

PKGH_224			
Training Method	Sampling Method	FID	KID
DDPM	DDPM	19.08	0.0134
	Epsilon Scaling(s=1.014)	39.88	0.0338
	DDIM	22.66	0.0154
LDM	DDPM	24.43	0.0185
	DDIM	25.56	0.0161

PKGH_336			
Training Method	Sampling Method	FID	KID
DDPM	DDPM	18.45	0.0129
	Epsilon Scaling(s=1.014)	45.12	0.0418
	DDIM	21.34	0.0159
LDM	DDPM	23.10	0.0160
	DDIM	26.41	0.0199

Table 4.3: FID and KID Scores for Different Models and Sampling Methods on Two Datasets

greater flexibility in capturing diverse tissue characteristics. These observations suggest that while both methods are effective, DDPM might be more reliable for producing uniformly detailed images, whereas Epsilon sampling can capture a broader range of textural variations. This could be a reason for the higher FID score of the epsilon scaling method.

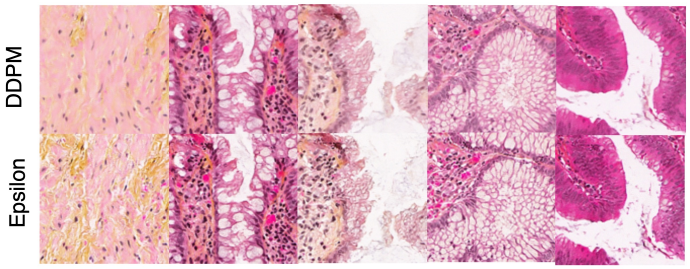


Figure 4.3: Comparison of generated pathology images using DDPM (top row) and Epsilon sampling (bottom row).

4.2.2 Analysis on Conditional Generation on pathology

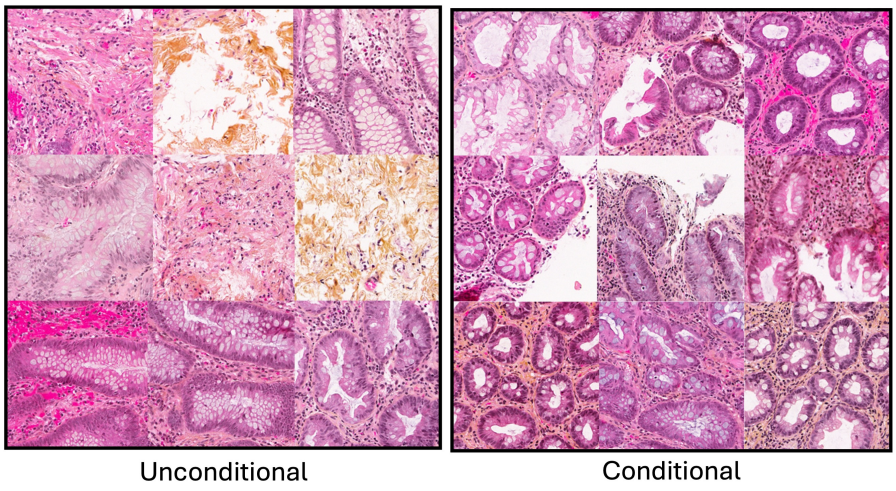
Dataset	$w = 0$	$w = 1$	$w = 2$	$w = 3$
PKGH_224	35.00	32.75	30.50	24.25
PKGH_336	28.47	25.10	23.89	21.02

Table 4.4: FID Scores for DDPM at Various Conditioning Levels for PKGH_224 and PKGH_336 Datasets (10k images used)

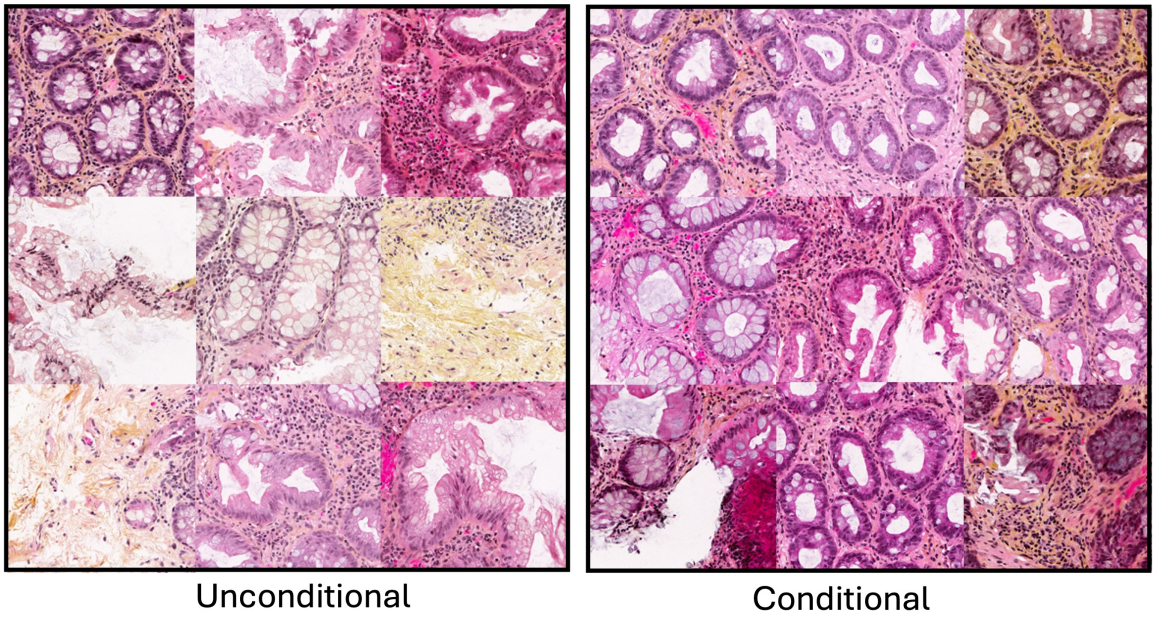
The table illustrates the positive impact of conditioning on FID scores and image realism for the PKGH_224 and PKGH_336 datasets. As the conditioning level increases from $w = 0$ to $w = 3$, FID scores consistently decrease for both datasets, indicating enhanced image quality. Specifically, PKGH_224's FID scores improve from 35.00 to 24.25, while PKGH_336's scores improve from 28.47 to 21.02. This trend demonstrates that higher conditioning levels enable the models to generate images with more accurate and detailed features, as seen in the more realistic and class-specific images produced by the conditional models in the Figure 4.4.

In the top row of Figure 4.4, the PKGH_224 dataset results show that the unconditional model (left panel) produces a variety of pathology images with broad structural characteristics. However, these images lack the specific detail necessary for a precise classification. In contrast, the conditional model (right panel), conditioned on the "Hyperplastic Polyp" class, generates images with high fidelity to the provided class. These images exhibit more accurate and detailed features characteristic of hyperplastic polyps, demonstrating the ability of the model to focus on and replicate specific class attributes. Similarly, the bottom row displays results from the PKGH_336 dataset.

By conditioning on a specific class, the model can produce high-quality, class-specific images that are essential for training robust machine learning models and advancing research in digital pathology. In summary, the conditional generation approach, using "Hyperplastic Polyp" as the class value, effectively demonstrates how generative diffusion models can enhance the quality and specificity of generated pathology images. This approach is pivotal in the development of accurate and detailed synthetic datasets.



(a) PKGH_224: Generated Pathology Images using Larger Fields of View (FOV).



(b) PKGH_336: Generated Pathology Images using Larger Fields of View (FOV).

Figure 4.4: Comparison of generated pathology images from different FOV datasets. The left panel shows pathology images generated using an unconditional diffusion model, while the right panel displays images generated using a conditional model(on the "Hyperplastic Polyp" class).

As shown in Figure 4.5, we applied different class-conditioning scales to DDPM. Each row in the figure represents an increase in the guidance scale, starting from $w = 0$ and progressing to $w = 3$. The results show a clear trend: As the guidance scale increases, the tissue structures generated become more distinct and defined. This highlights the critical role of the guidance scale

in enhancing the quality and clarity of synthetic pathology images.

For our visual analysis, our aim was to observe the effects of starting with the same initial noise across all varying the guidance scale. This approach allows us to assess how these factors influence the quality and characteristics of the generated images.

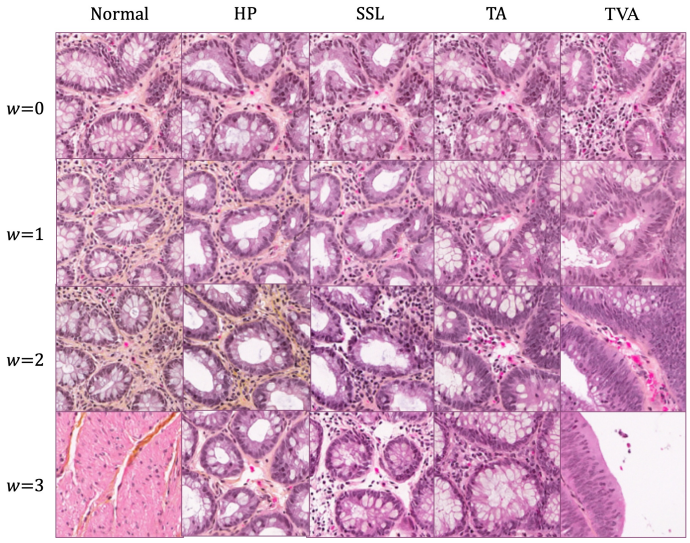


Figure 4.5: DDPM (Denoising Diffusion Probabilistic Models) with various class-conditioning: each row represents an increase in the guidance scale, starting from $w = 0$ and progressing to $w = 3$. The images reveal a clear trend: as guidance increases, the generated tissue structures become progressively more distinct and defined.

Furthermore, the diffusion generation process adapts differently according to the class embedding, demonstrating how much class information influences the generation based on the guidance scale. This analysis provides valuable insights on how the diffusion model is also separately for each class, allowing a better understanding of the different morphological patterns and the clear distinctions among tissue types. As we can see in the figure 4.5 even starting from the same initial noise, it generates different images, but structure and color look more similar to each other.

4.3 Analysis on Patch Size

The results indicate that the original patch size with a FOV of 224 and a patch resolution of 1.75 mpp achieved the lowest FID score of 19.08, reflecting the highest fidelity in generated images, which is expected as the images were trained on this model. However, examining the FID scores

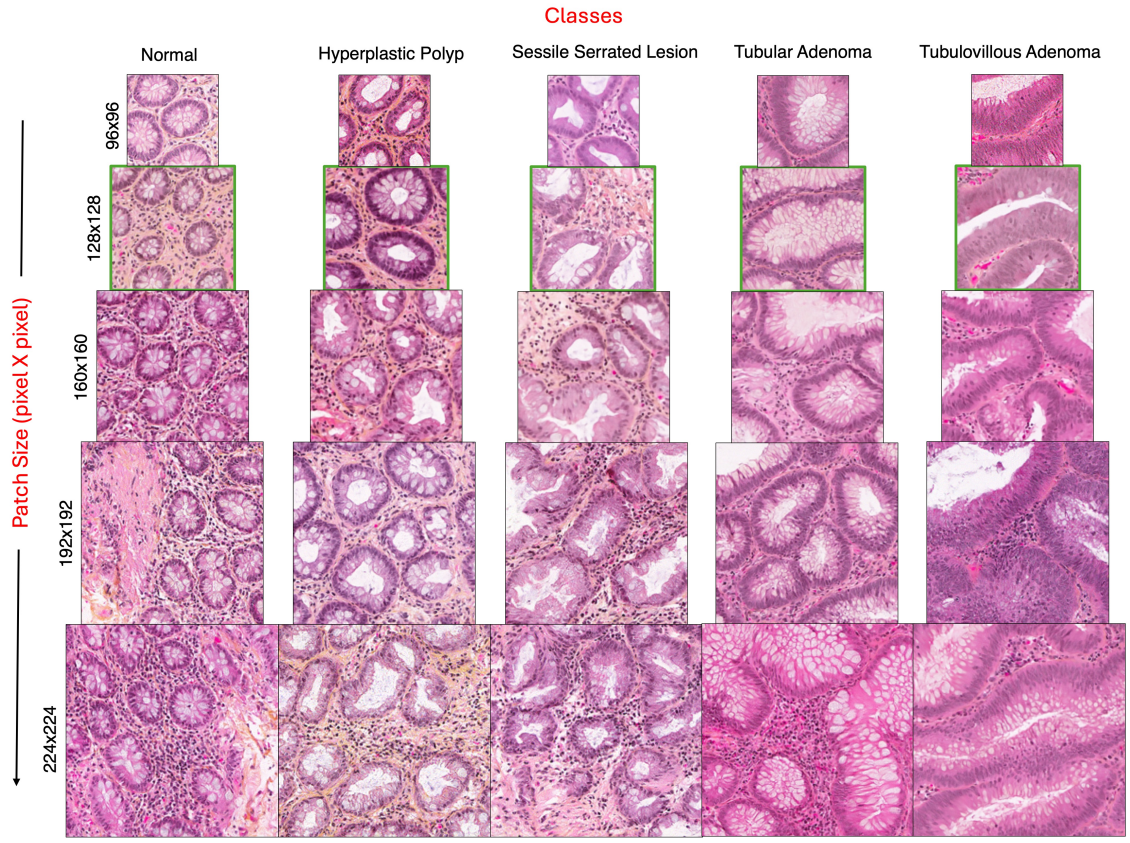


Figure 4.6: Patches generated for different patch size resembling different cell structure for five classes(128x128 is the training patch size used for diffusion model training)

for other patch sizes, especially those not used during training, provides insights into the model's ability to generate detailed images from different patch sizes.

Input	FoV patch (μm)	Patch reshaped	FID
Original Patch	224	128 × 128	**19.08**
Small Patch	112	64 × 64	161.01
	168	96 × 96	33.71
Large Patch	280	160 × 160	25.71
	336	192 × 192	38.41
	392	224 × 224	41.37

Table 4.5: Comparison of FID scores across different FOV and patch size.

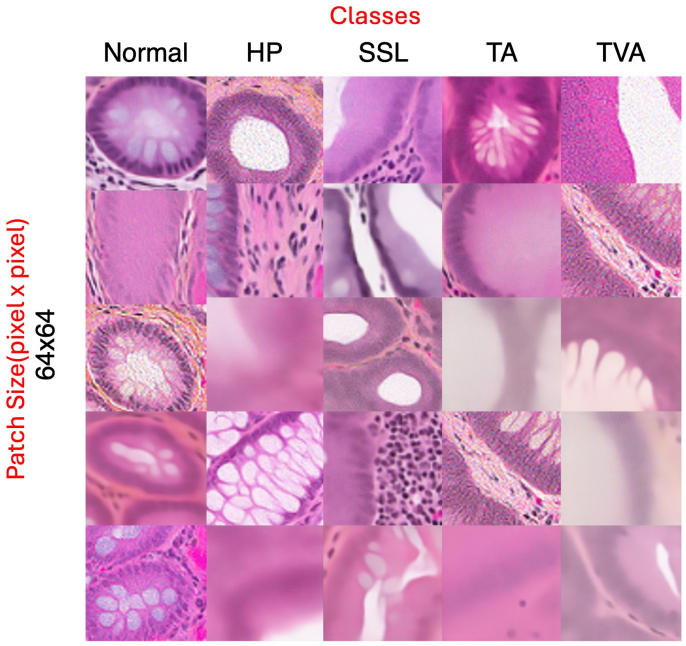


Figure 4.7: Patches generated for Patch size(64x64) which has the highest FID score

Our model is trained with a patch size of 128, and we experimented with generating images using different patch sizes, both smaller and larger. The table highlights the variation in image fidelity across different patch sizes and FOVs in generated pathology images. Smaller patch sizes, such as 64x64 and 96x96, exhibit higher FID scores of 161.01 and 33.71 respectively, indicating lower image fidelity. As the patch size increases to 160x160 and 192x192, the FID scores improve to 25.71 and 38.41 respectively, suggesting better image quality but with some variability. The 224x224 patch size, while larger, shows a slight decrease in fidelity with an FID score of 41.37. The visual representations in Figure 4.6 demonstrate that larger patch sizes generally capture more

detailed and high-level cellular structures across various classes, enhancing the potential utility of synthetic data in diagnostic workflows, especially when real data is limited. This indicates that while larger patches can improve detail and quality, the specific patch size must be carefully chosen to balance fidelity and practical application in diagnostic tasks. The FID score could vary as the model is not trained on these pixel levels and lacks the variability to accurately represent them.

Interestingly, even without training on certain pixel levels, the model can still perform well and generate quality images, demonstrating its robustness and versatility. This suggests that the generative model has the capacity to generalize beyond its training data to some extent, making it a valuable tool in scenarios where training data are scarce or diverse. This experiment highlights the model's adaptability and versatility, even when trained with data of limited complexity.

4.4 Evaluation Study of Pathology-Generated Images

Evaluation of generated pathology images is crucial to determine their quality and similarity to real pathology images. Although evaluation metrics provide valuable information about the quality of generated images, their applicability to real datasets remains uncertain. Therefore, this study aims to find the applicability of our generated dataset by evaluating its classification performance on the ResNet-50 architecture.

4.4.1 Methodology

To assess the quality of our generated images, we first trained the ResNet-50 model on the original dataset used to train the diffusion model. We reserved 20% of the WSIs as a test set to evaluate the accuracy of the model.

We applied a set of augmentations to enhance the robustness of the model, including random horizontal and vertical flipping, as well as transformations like jittering. For simplification, we used the ImageNet pretrained weights.

Our approach was not optimized for maximum performance, but was designed to compare the original and diffusion-generated models under conditional settings. The network was trained for 25 epochs with a batch size of 32. Given that the network utilized pre-trained ImageNet weights, the

initial learning rate was set to 0.0001, which was subsequently decreased using a cosine annealing scheduler.

We used the same settings to train the model on the generated images and tested it on the same test set. The performance of both models was evaluated using multiclass accuracy (ACC).

4.4.2 Results

The table below summarizes the classification performance on the KGH dataset across different scenarios, providing a clear comparison of the impact of real and generated data on model accuracy.

Dataset	**ACC (%)**
Real Dataset (PKGH_224)	89.95
Generated Dataset (PKGH_224)	88.62
Real + Generated (PKGH_224)	90.75
Real Dataset (PKGH_336)	94.06
Generated Dataset (PKGH_336)	92.44
Real + Generated (PKGH_336)	90.76

Table 4.6: Classification accuracy summary for PKGH_224 and PKGH_336: Classification accuracy scores of a ResNet-50 on synthetic data. Higher ACC proves the effectiveness of DGM-generated(DDPM) synthetic samples in capturing significant features.

The results demonstrate distinct differences in model accuracy when trained on real versus generated datasets, particularly between the PKGH_224 and PKGH_336 datasets. For the PKGH_224 dataset, the ResNet-50 model achieved an accuracy of 89.95% on the real dataset. However, the accuracy slightly decreased to 88.62% when trained solely on the generated dataset. Notably, when the real dataset was augmented with the generated data, the model's accuracy improved to 90.75%.

For the PKGH_336 dataset, the model exhibited even better performance, achieving an accuracy of 94.06% on the real dataset. The generated dataset alone resulted in an accuracy of 92.44%, and the combination of real and generated data yielded a slightly lower accuracy of 90.76%.

These findings indicate that while generated data alone can approach the performance of real data, combining real and generated data often enhances model accuracy, particularly for the PKGH_224 dataset. The consistently higher performance of the PKGH_336 dataset suggests that it provides more robust training data, possibly due to its larger FOV. This larger FOV likely captures more

contextual information and intricate details, allowing the model to learn more effectively from the data.

Overall, these results underscore the potential of using synthetic data to augment real datasets, especially when real data is limited. The improvements in classification accuracy highlight the effectiveness of Diffusion Generative Model (DGM)-generated synthetic samples in capturing essential features, thereby enhancing the model's ability to generalize from the data.

4.4.3 Conclusion

The results demonstrate that while generated data alone may not always surpass the performance of real data, it becomes highly valuable when combined with original datasets. The effectiveness of generated data is influenced by the dataset's complexity, particularly the FOV. In this study, higher FOV data, such as PKGH_336, even when generated synthetically, can approach or exceed the classification performance of real datasets. The combined use of both real and synthetic data significantly improves classification accuracy, making it a viable strategy to enhance model performance, especially in scenarios where real data are limited. This study highlights the potential of synthetic data as a valuable resource in the diagnostic workflow, particularly when available real data are scarce. Ongoing research to refine data generation techniques and their integration with real datasets will be crucial in achieving even better classification outcomes. These findings reinforce the utility of synthetic data in medical imaging, offering a practical solution for enhancing diagnosis workflows when data is limited.

4.5 Discussion

In the field of digital pathology, data is inherently complex, presenting challenges in comprehensive analysis and dataset generation that closely resembles original data. Our approach distinguishes itself by providing deeper insights into the use of Diffusion Generative Models (DGMs) and their capability to generate multiple patch sizes. Extensive experiments and evaluations were conducted to understand how these methods perform on intricate datasets.

The classification accuracy results reveal a noticeable difference when comparing models trained

on real datasets with those trained on generated datasets. While performance may decrease when only generated data are used with smaller FOV datasets, significant improvements are observed when real and synthetic datasets are combined. However, in the case of a more complex pathology dataset, such as PKGH_336, the highest accuracy is achieved using the generated dataset alone. This underscores the potential benefits of using synthetic data to enhance model performance, particularly when real data is scarce or difficult to obtain. The consistent improvement in accuracy with combined datasets (real + synthetic) demonstrates the effectiveness of DGM-generated synthetic samples in capturing significant features. This finding is crucial as it highlights how synthetic data can be a valuable supplement to real data, enhancing the robustness and accuracy of machine learning models.

In one study, we randomly select patches from specific fields of view (FOV) to compare the performance of the generated patches. Our findings indicate that DGMs offer a promising approach for new pathology data, as they can generate complex distributions that mimic the original data set while also uncovering hidden intricacies and predicting the structural details of cells and tissues across unseen image levels. This multilevel training approach highlights how effectively DGMs learn patch resolution.

Our results show that the FID scores of the generated patches are not exactly comparable to the original patch sizes used for training. Despite this, they provide a clear trend and a future direction for generating entire datasets by training on a limited number of image regions. The ability to generate complex, high-quality datasets from limited training data is particularly valuable in pathology, where obtaining comprehensive datasets can be challenging.

Comparison of the performance of our DDPM method with other studies demonstrates its strong performance in generating high-quality images. Our method outperforms some existing approaches and shows results comparable to others, indicating its competitive edge in the field. The lower FID scores achieved in our study highlight the importance of conditioning in DDPM and the significant impact of the dataset used.

Chapter 5

Conclusion and Future Work

In conclusion, this research successfully addressed key challenges in histopathology such as the need for high-quality synthetic datasets, demonstrating the practical applications of diffusion generative models (DGM) in medical imaging. The other unique primary experiment conducted in this thesis established that DGMs are capable of learning patch resolution effectively. By utilizing different patch sizes from the original dataset, we validated our findings. Although a direct comparison between Denoising Diffusion Probabilistic Models (DDPM) and Latent Diffusion Models (LDM) is challenging due to their different architectures, both models perform similarly well, leveraging the strengths of their respective designs. Among the various methods evaluated, DDPM proved to be highly flexible for working in pixel space, enabling the generation of detailed and realistic images. Meanwhile, LDM excelled in generating high-quality images even with larger models by using fewer time steps. Both models produced realistic images, as validated by their FID scores and visual analysis.

The research also addressed critical gaps using two datasets extracted from the WSIs of the Kingston General Hospital (KGH) dataset. By directly extracting data from the WSI, we were able to compare multiple FOV values. This approach allowed us to carefully select the FOVs that provided a balance between the number of images available for diffusion training and the classification test set. Our comparison of different FOVs from the same dataset revealed that larger FOVs not only yielded better FID scores, but also provided the highest classification accuracy. This selection process ensured that we maximized the effectiveness of our diffusion model training while maintaining

a robust test set for validation. We were able to enhance the overall performance and reliability of our generative models by utilizing larger FOVs that encompass more contextual information and tissue structures that can contribute to a richer representation of the underlying biological features. This comprehensive data representation improves the model's ability to learn and generalize, leading to superior performance in both generative and classification tasks.

In general, this thesis highlights the meaningful findings in the field of computational pathology by providing a novel and comprehensive study of DGM and sets the stage for future enhancements. The thesis presents practical applications and addresses key challenges, paving the way for future research and development. The ability of DGMs to generate realistic and high-quality synthetic pathology images, especially with larger FOVs, highlights their potential to enhance the robustness and accuracy of deep learning classifiers.

5.1 Future Work

Future research will focus on several key areas to further advance the application of diffusion models in medical imaging.

Biomarker Identification and Analysis: Future work will explore the potential of diffusion models to identify and analyze biomarkers for various diseases (Echle et al., 2021). This involves developing methodologies to enhance the accuracy and reliability of biomarker discovery, leveraging the advanced capabilities of diffusion models to generate new subtle and critical features in medical images that are indicative of specific biomarkers.

Synthetic Cross-Staining Images for Biomarker Studies: To further improve biomarker studies, synthetic cross-stained images can be created, such as converting Hematoxylin and Eosin (H&E) stained images (Boschman et al., 2022) to Immunohistochemistry (IHC) stained images (Rojo, Bueno, & Slodkowska, 2009). These synthetic images can be evaluated for their effectiveness in providing additional insights and enhancing the understanding of biomarkers which is why IHC are famous. Cross-staining synthesis is particularly helpful when other staining techniques, like IHC, are beneficial but expensive. This approach is expected to offer new perspectives and deepen the analysis of biomarker characteristics, potentially revealing patterns and details that are not apparent

with a single staining method.

Impact of Magnification Levels on Diffusion Models: The research can be done to determine the optimal magnification settings for accurate image analysis and synthesis in histopathology. This research will assess how different magnification levels affect the performance of diffusion models, aiming to identify the settings that produce the highest quality image.

Multi-Class Labeling Techniques for Histopathology Images: Future research can implement multi-class labeling techniques to enhance the classification of histopathology images. This involves conditioning diffusion models on multi-class labels to achieve more precise image analysis and interpretation. The aim is to improve the accuracy and robustness of classifiers in medical imaging, leading to better diagnostic and analytical outcomes.

These research directions aim to expand the current understanding and application of diffusion models in medical imaging and biomarker discovery. They will contribute to developing more accurate and privacy-sensitive healthcare solutions.

Appendix A

Unified Psuedo Code for Diffusion Generative Models

The Unified-Pseudo Code algorithm provides a comprehensive framework for training and sampling in Diffusion Generative Models (DGMs) mention in Chapter 2. Training and sampling processes are clearly delineated within the code, highlighting their distinct roles and methodologies. The training phase, contained in the `TRAINING(y)` function, focuses on iteratively learning from data through the `DiffusionProcess(y, t)` function. This involves generating noisy versions of the data and refining them to minimize the loss, thereby enabling the model to learn patch resolution effectively. The `DiffusionProcess` function adds noise to the data during training to perturb the input and enhance the model's robustness.

Conversely, the sampling phase, detailed in the `SAMPLING(`ϵ_θ`)` function, aims to generate new data samples from the learned model. This process begins with instantiating random noise and iteratively refining it using various sampling techniques like Epsilon Scaling, Adaptive Momentum Sampling, and ERA-Solver Sampling, depending on the specific sampling method being employed. Each sampling method modifies the noise reduction steps uniquely: Epsilon Scaling adjusts noise proportionally, Adaptive Momentum Sampling incorporates momentum to speed up convergence, and ERA-Solver Sampling uses an error-regularized approach for precise noise correction. The distinction between methods like DDPM and DDIM is also emphasized, with DDPM leveraging direct

denoising steps and DDIM using deterministic sampling to improve image quality and sampling efficiency.

Algorithm 1 Unified-Pseudo Code

Training: Training Algorithm(1.1 DDPM, 1.2 Input-Perturbation)
Sampling: Sampling Algorithm (2.1 DDPM, 2.2 DDIM, 2.3 Epsilon Scaling, 2.4 Adaptive Momentum, 2.5 ERA-Solver Sampling,)
Domain: Image domain (3.1 Pixel, 3.2 Latent, 3.3 Gradient)
E($\cdot$): Encoder Network
D($\cdot$): Decoder Network
γ: Noise schedule for ξ
$\bar{\alpha}$: Eq. 2
function DIFFUSIONPROCESS(y, t)**:**
    $\epsilon \leftarrow$ torch.randn(y)
    if Training is Input-Perturbation **then** $\epsilon = \epsilon + \gamma\xi$ # 1.2 Input Perturbation Training
    end if
    $\tilde{y} \leftarrow \sqrt{\bar{\alpha}_t}\, y + \sqrt{1-\bar{\alpha}_t} * \epsilon$ # 1.1 DDPM Training
    return $\tilde{y}, \epsilon$
end function
function TRAINING(y)**:**
    for t **in** (1,...,T) **do**
        $\tilde{y}, \epsilon_t \leftarrow$ DiffusionProcess(y, t)
        $\bar{\epsilon}_t \leftarrow \epsilon_\theta(\tilde{y}, t)$
        loss $\leftarrow$ L$(\bar{\epsilon}_t, \epsilon_t)$
    end for
end function

function SAMPLING(ϵ_θ)**:**
    Instantiate : $x_T \sim \mathcal{N}(0, I)$
    for t **in** (T,...,1) **do**
        z $\leftarrow$ torch.randn_like(x) **if** $t > 1$, **else** z $= 0$
        $\epsilon_t \leftarrow \epsilon_\theta(x, t)$
        if Sampling is Epsilon Scaling **then** # 2.3 Epsilon Scaling
            $\epsilon_t = \frac{\epsilon_t}{\lambda_t}$ Eq. 22
        end if
        if Sampling is Adaptive Momentum Sampling **then** # 2.4 Adaptive Momentum
            $x_t \leftarrow \left(\sqrt{\frac{1-\alpha_{t-1}-\sigma_t^2}{\alpha_{t-1}}} - \sqrt{\frac{1-\alpha_t}{\alpha_t}}\right) \cdot \epsilon_\theta^{(t)}(x_t) + \frac{\sigma_t}{\sqrt{\alpha_{t-1}}} \cdot \epsilon_t$
            $\mathbf{v_{t-1}} = (1-c) \cdot v_t + c \cdot \|\mathbf{d\bar{x}_t}\|^2$

            $\mathbf{m_{t-1}} = a \cdot \mathbf{m_t} + b \cdot (\mu \cdot \epsilon_\theta^{(t)}(x_t) + \frac{\sigma_t}{\sqrt{\alpha_{t-1}}} \cdot \epsilon_t)$
            $x_{t-1} = \mathbf{\bar{x}_t} + \frac{\mathbf{m_{t-1}}}{\sqrt{\mathbf{v_{t-1}}}+\zeta}$
        end if
    end for
    return x_0
end function

Algorithm 1 Unified-Pseudo Code (continued)

if Sampling is ERA-Solver **then** # 2.5 ERA-Solver Sampling
  Input: k
  $i = t$
  if $i == T$ **then**
    Instantiate: $\Delta\epsilon = \lambda, \Omega = \Omega \cup \{(t_i, \epsilon_\theta(x_{t_i}, t_i))\}$,
  end if
  if $(T - i) < k - 1$ **then**
    $\epsilon_{t_i} = \epsilon_\theta(x_{t_i}, t_i)$
  else
    $\{\bar{\tau}_m\}_{m=0}^{k-1} \leftarrow$ for each$\bar{\tau}_m = \frac{i}{k} \cdot m$
    $\{\tau_m\}_{m=0}^{k-1} \leftarrow$ for each$\tau_m = [\left(\frac{\bar{\tau}_m}{i}\right)^{\Delta\varepsilon/\lambda} \cdot i]$
    $\bar{\epsilon}_\theta(x_{t_i-1}, t_{i-1}) \leftarrow L_\epsilon(t_{i-1})$.
    $\epsilon_{t_i} = \frac{1}{24}\left(9\bar{\epsilon}_\theta(x_{t_i-1}, t_{i-1}) + 19\epsilon_\theta(x_{t_i}, t_i) - 5\epsilon_\theta(x_{t_{i+1}}, t_{i+1}) + \epsilon_\theta(x_{t_{i+2}}, t_{i+2})\right)$
    $\Delta\epsilon = \|\epsilon_\theta(x_{t_i}, t_i) - \bar{\epsilon}_\theta(x_{t_i}, t_i)\|_2$
  end if
  $\Omega = \Omega \cup \{(t_{i-1}, \epsilon_\theta(x_{t_{i-1}}, t_{i-1}))\}$
  $\epsilon_t = \epsilon_{t_i}$
end if
if Sampling is DDPM **then**
  $x_{t-1} \leftarrow \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}} \cdot \epsilon_t + \sqrt{\beta_t} \cdot \mathrm{z}\right)$ # 2.1 DDPM Sampling
else if Sampling is DDIM **then**
  $x_{t-1} = \sqrt{\alpha_{t-1}}\left(\frac{x_t - \sqrt{1-\alpha_t} \cdot \epsilon_t}{\sqrt{\alpha_t}}\right) + \sqrt{1 - \alpha_{t-1} - \sigma_t^2} \cdot \epsilon_t + \sigma_t \cdot \mathrm{z}$ # 2.2 DDIM Sampling
end if
# N: No of samples to be generated

# Training
for x **in**(dataloader) **do**
  if domain is pixel **then** $y = x$ # 3.1 Pixel Domain
  else if domain is latent **then** $y = E(x)$ # 3.2 Latent Domain
  else if domain is gradient **then** $y = \nabla x$ # 3.3 Gradient Domain
  end if
  ϵ_θ = Training(x)
end for

# Sampling
for (i in 0...N) **do**
  y_0 = Sampling(ϵ_θ)
  if domain is pixel **then** $x_0 = y_0$ # 3.1 Pixel Domain
  else if domain is latent **then** $x = D(y_0)$ # 3.2 Latent Domain
  else if domain is gradient **then** $x =$ PoissonEquation(y_0) # 3.3 Gradient Domain
  end if
end for

References

Basavanhally, A., Ganesan, S., Shih, N., Mies, C., Feldman, M., Tomaszewski, J., & Madabhushi, A. (2011). A boosted classifier for integrating multiple fields of view: Breast cancer grading in histopathology. In *2011 ieee international symposium on biomedical imaging: From nano to macro* (pp. 125–128).

Baxi, V., Edwards, R., Montalto, M., & Saha, S. (2022). Digital pathology and artificial intelligence in translational medicine and clinical practice. *Modern Pathology*, *35*(1), 23–32.

Bińkowski, M., Sutherland, D. J., Arbel, M., & Gretton, A. (2018). Demystifying mmd gans. *arXiv preprint arXiv:1801.01401*.

Boschman, J., Farahani, H., Darbandsari, A., Ahmadvand, P., Van Spankeren, A., Farnell, D., ... others (2022). The utility of color normalization for ai-based diagnosis of hematoxylin and eosin-stained pathology images. *The Journal of Pathology*, *256*(1), 15–24.

Carlini, N., Hayes, J., Nasr, M., Jagielski, M., Sehwag, V., Tramer, F., ... Wallace, E. (2023). Extracting training data from diffusion models. In *32nd usenix security symposium (usenix security 23)* (pp. 5253–5270).

Chen, R. J., Lu, M. Y., Chen, T. Y., Williamson, D. F., & Mahmood, F. (2021). Synthetic data in machine learning for medicine and healthcare. *Nature Biomedical Engineering*, *5*(6), 493–497.

Cho, J., Lee, K., Shin, E., Choy, G., & Do, S. (2015). How much data is needed to train a medical image deep learning system to achieve necessary high accuracy? *arXiv preprint arXiv:1511.06348*.

Daniel, N., Aknin, E., Larey, A., Peretz, Y., Sela, G., Fisher, Y., & Savir, Y. (2023). Between

generating noise and generating images: Noise in the correct frequency improves the quality of synthetic histopathology images for digital pathology. In *2023 45th annual international conference of the ieee engineering in medicine biology society (embc)* (p. 1-7). doi: 10.1109/EMBC40787.2023.10341042

Dhariwal, P., & Nichol, A. (2021). Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, *34*, 8780–8794.

Echle, A., Rindtorff, N. T., Brinker, T. J., Luedde, T., Pearson, A. T., & Kather, J. N. (2021). Deep learning in cancer pathology: a new generation of clinical biomarkers. *British journal of cancer*, *124*(4), 686–696.

Falahkheirkhah, K., Tiwari, S., Yeh, K., Gupta, S., Herrera-Hernandez, L., McCarthy, M. R., ... Bhargava, R. (2023). Deepfake histologic images for enhancing digital pathology. *Laboratory Investigation*, *103*(1), 100006.

Gong, Y. (2023). Gradient domain diffusion models for image synthesis. *arXiv preprint arXiv:2309.01875*.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, *63*(11), 139–144.

Graikos, A., Yellapragada, S., Le, M.-Q., Kapse, S., Prasanna, P., Saltz, J., & Samaras, D. (2024). Learned representation-guided diffusion models for large-image generation. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition* (pp. 8532–8542).

Han, X., Zheng, H., & Zhou, M. (2022). Card: Classification and regression diffusion models. *Advances in Neural Information Processing Systems*, *35*, 18100–18115.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 770–778).

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. (2017). Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, *30*.

Ho, J., Chan, W., Saharia, C., Whang, J., Gao, R., Gritsenko, A., ... others (2022). Imagen video: High definition video generation with diffusion models. *arXiv preprint arXiv:2210.02303*.

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in neural information processing systems*, *33*, 6840–6851.

Ho, J., Saharia, C., Chan, W., Fleet, D. J., Norouzi, M., & Salimans, T. (2022). Cascaded diffusion models for high fidelity image generation. *Journal of Machine Learning Research*, *23*(47), 1–33.

Ho, J., & Salimans, T. (2022). Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*.

Hosseini, M. S., Bejnordi, B. E., Trinh, V. Q.-H., Chan, L., Hasan, D., Li, X., ... others (2024). Computational pathology: a survey review and the way forward. *Journal of Pathology Informatics*, 100357.

Jahn, S. W., Plass, M., & Moinfar, F. (2020). Digital pathology: advantages, limitations and emerging perspectives. *Journal of clinical medicine*, *9*(11), 3697.

Janowczyk, A., & Madabhushi, A. (2016). Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. *Journal of pathology informatics*, *7*(1), 29.

Jaravaza, D. R., & Rigby, J. M. (2020). Hyperplastic polyp or sessile serrated lesion? the contribution of serial sections to reclassification. *Diagnostic pathology*, *15*, 1–9.

Jass, J., & Burt, R. (2000). Hyperplastic polyposis. *Pathology and genetics of tumours of the digestive system*, 135.

Kapil, A., Meier, A., Zuraw, A., Steele, K. E., Rebelatto, M. C., Schmidt, G., & Brieu, N. (2018). Deep semi supervised generative learning for automated tumor proportion scoring on nsclc tissue needle biopsies. *Scientific reports*, *8*(1), 17343.

Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.

Kiran, N., Sapna, F., Kiran, F., Kumar, D., Raja, F., Shiwlani, S., ... others (2023). Digital pathology: transforming diagnosis in the digital age. *Cureus*, *15*(9).

Kuklyte, J., Fitzgerald, J., Nelissen, S., Wei, H., Whelan, A., Power, A., ... others (2021). Evaluation of the use of single-and multi-magnification convolutional neural networks for the determination and quantitation of lesions in nonclinical pathology studies. *Toxicologic Pathology*, *49*(4), 815–842.

Levine, A. B., Peng, J., Farnell, D., Nursey, M., Wang, Y., Naso, J. R., ... others (2020). Synthesis of diagnostic quality cancer pathology images by generative adversarial networks. *The Journal of pathology*, *252*(2), 178–188.

Li, H., Yang, Y., Chang, M., Chen, S., Feng, H., Xu, Z., ... Chen, Y. (2022). Srdiff: Single image super-resolution with diffusion probabilistic models. *Neurocomputing*, *479*, 47–59.

Li, S., Liu, L., Chai, Z., Li, R., & Tan, X. (2023). Era-solver: Error-robust adams solver for fast sampling of diffusion probabilistic models. *arXiv preprint arXiv:2301.12935*.

Moghadam, P. A., Van Dalen, S., Martin, K. C., Lennerz, J., Yip, S., Farahani, H., & Bashashati, A. (2023). A morphology focused diffusion probabilistic model for synthesis of histopathology images. In *Proceedings of the ieee/cvf winter conference on applications of computer vision* (pp. 2000–2009).

Murakami, T., Kurosawa, T., Fukushima, H., Shibuya, T., Yao, T., & Nagahara, A. (2022). Sessile serrated lesions: Clinicopathological characteristics, endoscopic diagnosis, and management. *Digestive Endoscopy*, *34*(6), 1096–1109.

Nichol, A. Q., & Dhariwal, P. (2021). Improved denoising diffusion probabilistic models. In *International conference on machine learning* (pp. 8162–8171).

Ning, M., Li, M., Su, J., Salah, A. A., & Ertugrul, I. O. (2023). Elucidating the exposure bias in diffusion models. *arXiv preprint arXiv:2308.15321*.

Ning, M., Sangineto, E., Porrello, A., Calderara, S., & Cucchiara, R. (2023). Input perturbation reduces exposure bias in diffusion models. *arXiv preprint arXiv:2301.11706*.

O'connor, J. P., Aboagye, E. O., Adams, J. E., Aerts, H. J., Barrington, S. F., Beer, A. J., ... others (2017). Imaging biomarker roadmap for cancer studies. *Nature reviews Clinical oncology*, *14*(3), 169–186.

Oh, H.-J., & Jeong, W.-K. (2023). Diffmix: Diffusion model-based data synthesis for nuclei segmentation and classification in imbalanced pathology image datasets. In *International conference on medical image computing and computer-assisted intervention* (pp. 337–345).

Pocock, J., Graham, S., Vu, Q. D., Jahanifar, M., Deshpande, S., Hadjigeorghiou, G., ... Raza, S. E. A. (2022, sep). TIAToolbox as an end-to-end library for advanced tissue image analytics. *Communications Medicine*, *2*(1), 120. Retrieved from https://www.nature.com/

articles/s43856-022-00186-5 doi: 10.1038/s43856-022-00186-5

Pozzi, M., Noei, S., Robbi, E., Cima, L., Moroni, M., Munari, E., ... Jurman, G. (2023). Generating synthetic data in digital pathology through diffusion models: a multifaceted approach to evaluation. *medRxiv*, 2023–11.

Price, W. N., & Cohen, I. G. (2019). Privacy in the age of medical big data. *Nature medicine*, *25*(1), 37–43.

Qin, Y., Zheng, H., Yao, J., Zhou, M., & Zhang, Y. (2023). Class-balancing diffusion models. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition* (pp. 18434–18443).

Quiros, A. C., Murray-Smith, R., & Yuan, K. (2019). Pathologygan: Learning deep representations of cancer tissue. *arXiv preprint arXiv:1907.02644*.

Rojo, M. G., Bueno, G., & Slodkowska, J. (2009). Review of imaging solutions for integrated quantitative immunohistochemistry in the pathology daily practice. *Folia histochemica et cytobiologica*, *47*(3), 349–354.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition* (pp. 10684–10695).

Salemis, N. S., Gemenetzis, G., Karagkiouzis, G., Seretis, C., Sapounas, K., Tsantilas, V., ... Lagoudianakis, E. (2012). Tubular adenoma of the breast: a rare presentation and review of the literature. *Journal of Clinical Medicine Research*, *4*(1), 64.

Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., & Chen, X. (2016). Improved techniques for training gans. *Advances in neural information processing systems*, *29*.

Savage, N. (2023). Synthetic data could be better than real data. *Nature*.

Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., & Ganguli, S. (2015). Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning* (pp. 2256–2265).

Song, J., Meng, C., & Ermon, S. (2020). Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*.

Song, Y., & Ermon, S. (2019). Generative modeling by estimating gradients of the data distribution.

Advances in neural information processing systems, *32*.

Song, Y., & Ermon, S. (2020). Improved techniques for training score-based generative models. *Advances in neural information processing systems*, *33*, 12438–12448.

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., & Poole, B. (2020). Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... Rabinovich, A. (2015). Going deeper with convolutions. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 1–9).

Tang, B., Ewalt, J., & Ng, H.-L. (2021). Generative ai models for drug discovery. In *Biophysical and computational tools in drug discovery* (pp. 221–243). Springer.

Tang, Y., Tang, Y., Zhu, Y., Xiao, J., & Summers, R. M. (2021). A disentangled generative model for disease decomposition in chest x-rays via normal image synthesis. *Medical Image Analysis*, *67*, 101839.

Van Den Oord, A., Vinyals, O., et al. (2017). Neural discrete representation learning. *Advances in neural information processing systems*, *30*.

Van der Laak, J., Litjens, G., & Ciompi, F. (2021). Deep learning in histopathology: the path to the clinic. *Nature medicine*, *27*(5), 775–784.

Wang, T.-C., Liu, M.-Y., Zhu, J.-Y., Tao, A., Kautz, J., & Catanzaro, B. (2018). High-resolution image synthesis and semantic manipulation with conditional gans. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 8798–8807).

Wang, X., Dinh, A.-D., Liu, D., & Xu, C. (2023). Boosting diffusion models with an adaptive momentum sampler. *arXiv preprint arXiv:2308.11941*.

Wiatrak, M., Albrecht, S. V., & Nystrom, A. (2020, March 24). Stabilizing generative adversarial networks: A survey.
(Available at arXiv:2003.12697)

Xiao, Z., Kreis, K., & Vahdat, A. (2021). Tackling the generative learning trilemma with denoising diffusion gans. *arXiv preprint arXiv:2112.07804*.

Yellapragada, S., Graikos, A., Prasanna, P., Kurc, T., Saltz, J., & Samaras, D. (2024). Pathldm: Text

conditioned latent diffusion model for histopathology. In *Proceedings of the ieee/cvf winter conference on applications of computer vision* (pp. 5182–5191).

Zhou, T., Fu, H., Chen, G., Shen, J., & Shao, L. (2020). Hi-net: hybrid-fusion network for multi-modal mr image synthesis. *IEEE transactions on medical imaging*, *39*(9), 2772–2781.