

A Morphology Focused Diffusion Probabilistic Model for Synthesis of Histopathology Images

Puria Azadi Moghadam *
University of British Columbia
puria.azadi@ubc.ca

Jochen Lennerz
Harvard Medical School
jlennerz@partners.org

Sanne Van Dalen *
Eindhoven University of Technology
s.l.v.dalen@student.tue.nl

Stephen Yip
BC Cancer Agency
SYip-02@bccancer.bc.ca

Karina C. Martin
BC Cancer Agency
karina.choronenka@vch.ca

Hossein Farahani
University of British Columbia
h.farahani@ubc.ca

Ali Bashashati
University of British Columbia
ali.bashashati@ubc.ca

Abstract

Visual microscopic study of diseased tissue by pathologists has been the cornerstone for cancer diagnosis and prognostication for more than a century. Recently, deep learning methods have made significant advances in the analysis and classification of tissue images. However, there has been limited work on the utility of such models in generating histopathology images. These synthetic images have several applications in pathology including utilities in education, proficiency testing, privacy, and data sharing. Recently, diffusion probabilistic models were introduced to generate high quality images. Here, for the first time, we investigate the potential use of such models along with prioritized morphology weighting and color normalization to synthesize high quality histopathology images of brain cancer. Our detailed results show that diffusion probabilistic models are capable of synthesizing a wide range of histopathology images and have superior performance compared to generative adversarial networks.

1. Introduction

Histopathology is a diagnostic science that relies on the visual examination of cellular and tissue characteristics in magnified tissue slides[17]. Recently, high-throughput digital pathology scanners have been developed that can provide gigapixel high-resolution images($\sim 100K \times 100K$ pixels) of microscope slides at objective magnifications of up to $40\times$. Furthermore, histological staining of tissues with

various stains (e.g., hematoxylin and eosin, silver nitrate, carmine, hematin, etc.) is used to emphasise the properties of the tissues and improve their contrast for examination [1]. Figure 1 shows a sample of digital pathology images.

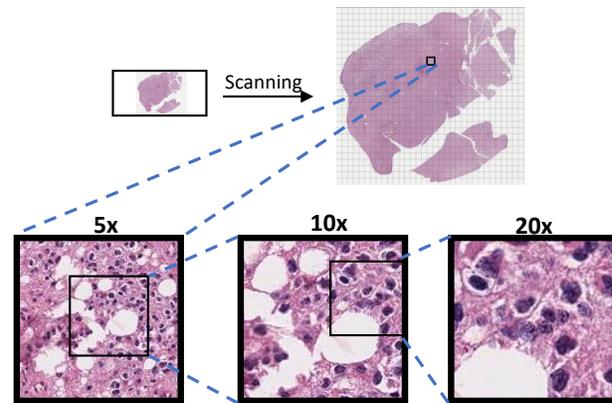


Figure 1: 3 sampled patches at different magnifications from a whole slide image

The most commonly used stain material is Hematoxylin and Eosin (H&E), which stains the nucleic acid within the cell nuclei with purplish blue and extracellular matrix and cytoplasm by pink color. Afterwards, pathologists examine the cytological and tissue characteristics of the sample for cancer diagnosis and staging. The histopathological diagnosis of cancer is time-consuming and is prone to subjective differences[44], as it is heavily reliant on pathologists' experiences and prior exposure to various histological variants (i.e., subtypes). However, some of these variants are

*Equal Contribution

rare, and pathologists do not get the opportunity to examine them during their training. In addition, there is a scarcity of samples for certain subtypes in various parts around the world.

Following the success of deep learning models in the past decade in other applications, various models have been developed in recent years to improve the quality of histopathology diagnosis and assist pathologists in the decision making process. However, majority of the research efforts have been on developing discriminative models with more emphasis on classification tasks [37]. Wang et al. [41], for example, employed an accelerated deep convolution neural network for both cell identification and classification. Mathur et al. [27] developed a multi-gaze attention network with multiheaded self-attention for renal biopsy image categorization tasks. Segmentation is another frequent application of deep models in histopathology. The authors in [15] presented their multi-resolution UNet-based approach for detecting breast cancer metastasis. Another example is the use of a deep learning algorithm for the prediction of patients' outcomes[29].

Compared to discriminative models, utilizing generative models in pathology is still in its infancy. Generative models can be used to create synthetic images and have numerous potential applications in pathology and can entail discovering patterns and regularities that may give more information about distinctions across subtypes.

Histology generative models have a significant educational value and they can also be utilized to synthesize histology images for proficiency testing applications. [23]. Specially for some rare cancers there is a shortage of diverse examples for training pathologists. Histology generative models are able to provide a limitless number of new images for teaching and examination in those cancers. In addition, these images can tackle the privacy concerns of sharing pathology images, which is one of the emerging challenges with patient privacy and data protection[8]. Although, unlike genomics data, it is not possible to identify patients from their histopathology slides, still due to very strict data privacy regulations in healthcare, sharing the original histopathology slides between institutions is extremely difficult. Histology generative models are the best way for overcoming such restrictions. There are a few distributed-based learning methods for histopathology that have recently been developed [25] and Rajotte et al. [34] showed an approach for addressing this problem is to train generative models on local private data and only share only the synthetic data or the generative model itself.

In comparison to normal images, several factors, such as time, labor, and economic costs, make having fully labeled medical images more difficult [22], causing the model to suffer from overfitting. Synthesized histology images can be used to augment the datasets and improve the perfor-

mance of trained models. Taken together, they have the potential to improve the educational, assessment, security, and generalization aspects of the histology field.

Generative Adversarial Networks (GAN) introduced by Goodfellow et al.[13] and its later variants are currently the most common models for generating synthetic images. However, GANs have several drawbacks, such as mode collapse and difficulties in training.

Starting from Ho et al.[16] several studies showed that diffusion probabilistic models can generate high-fidelity images comparable to those generated by GANs [32, 10, 7]. Diffusion probabilistic models offer several desirable properties for image synthesis, such as stable training, easy model scaling, and good distribution coverage. However, their performance has not been explored for histopathology images.

The goal of this paper is to explore the utility of diffusion probabilistic models for synthesizing histopathology images and comparing these models with state-of-the-art for quality of the generated images and lack of artefacts.

The contributions of this paper are as follows:

- In this paper, for the first time, we propose exploiting diffusion probabilistic models to generate synthetic histopathology images. We use genotype data as class labels to show the power of generative models in synthesizing subtle differences corresponding with genotypes in one single cancer type. This is a more challenging task compared to generating images of different cancer types. Nevertheless, genotype is not a requirement for the model and it can work with any data label.
- We also benefit from color normalization to force our end-to-end model to learn morphological patterns and from perception prioritized weighting (P2) [7], which aims to prioritize focusing on diffusion stages with more important structural histopathology contents.
- We conduct an extensive empirical study using a low grade glioma (LGG) dataset and compare the performance of the proposed generation method against a state-of-the-art study that utilized GANs for histopathology image analysis, using a variety of metrics. Our results show that the introduced method is superior in all of them, and it produces highly realistic pathology images. The used dataset and our results are described in section 4.

2. Related Works

Generating Histopathology images has been getting popular in recent years because of advances in digital pathology imaging and computational infrastructures as well as the in-

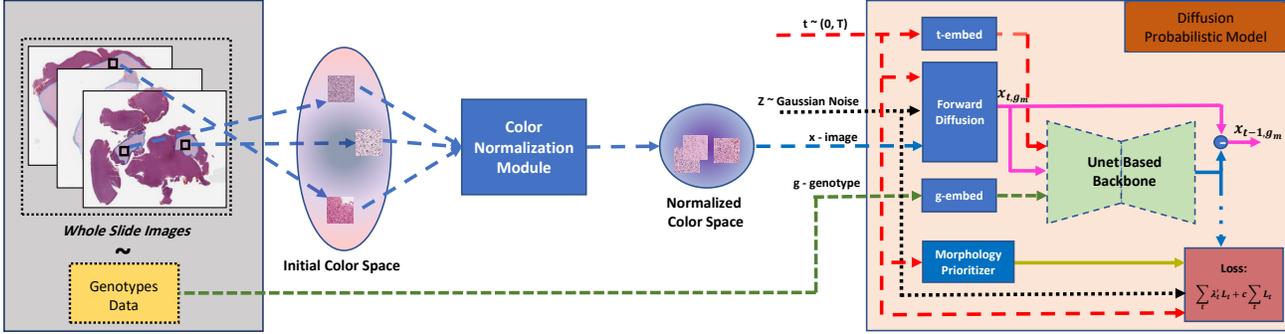


Figure 2: The overview of proposed approach for generating histopathology images with genotypes guidance

roduction of powerful deep generative models that are able to tackle specific difficulties in this domain.

In 2018, Kapil et al. [18] utilized generative learning to enable automated tumor proportion scoring. Zhou et al. [46] employed a U-net based GAN to augment and generate high-resolution image from low-resolution histology patches. More recently, the Progressive GAN model has shown considerable potential in generating high quality histopathological images [23, 45].

GAN models, on the other hand, suffer from mode collapse and instabilities due to directly producing images from complex latent spaces in a single shot and easily overfitting of their discriminator[43], which make them unsuitable for generating samples from rare conditions or imbalanced datasets.

However, diffusion probabilistic models were recently introduced and used for several applications. Amit et al.[2] used a diffusion model for performing image segmentation. Other examples are the use of diffusion probabilistic models for converting text to speech[33] and generating online handwriting[26]. As diffusion models divide the process into several relatively simple denoising diffusion steps and make strong conditioning on input images at each step, they soften the data distribution, which results in models with higher stability. These models also are able to generate more diverse images with better distribution coverage, are less probable to get overfitted, and are easy for scaling [43].

These advantages can significantly improve the histopathology image generation domain as imbalanced datasets with rare subtypes is one of the main issues in this domain. Although diffusion probabilistic models are used within different areas for multiple tasks, they are not exploited to generate synthetic histopathology images yet. Therefore, we hypothesize that such models could generate high quality histopathology images that may address some of the challenges of GAN-based models.

3. Method

3.1. Problem Definition

The objective of this paper is to enable the generation of histopathology images that are represented by various morphologic features. Synthesizing these pathology images is a challenging task compared to typical images in other domains.

Assume $g_{i=1,2,\dots,k}$ indicates the i -th available genotype. For each genotype, there is:

$$Set_{g_i} = \{image_j\}, j = 1, 2, \dots, N \ \& \ image_j \in \mathbb{R}^{c \times h \times w}, \quad (1)$$

which:

$$genotype(image_j) = g_i, \quad (2)$$

The purpose is to have an estimator function f_{est} that:

$$f_{est}\{g_i, n \sim Noise\} = image_{g_i, out}, \quad (3)$$

while:

$$image_{g_i, out} \in \mathbb{R}^{c \times h \times w}, \ image_{g_i, out} \sim D(Set_{g_i}), \quad (4)$$

where $D(Set_{g_i})$ is the distribution of the images associated with g_i . we tackle LGG, which account for the majority of pediatric brain tumors [35] and they are classified by combining the histopathological features with genotype features since 2016[24] including isocitrate dehydrogenase (IDH) and co-deletion of 1p19q which are the short arm of chromosome 1 and the long arm of chromosome 19[31, 6]. Diagnosis of LGG is done through histopathologic examination of tissue. Figure 2 summarizes our proposed end-to-end solution for generating histology images.

3.2. Color normalization

One of the main challenges related to H&E images is the lack of consistency in the staining due to variances in site-specific staining protocol or digital scanning platforms and methodologies. Color normalization strategies are able to boost histological discriminative models' performance.[4].

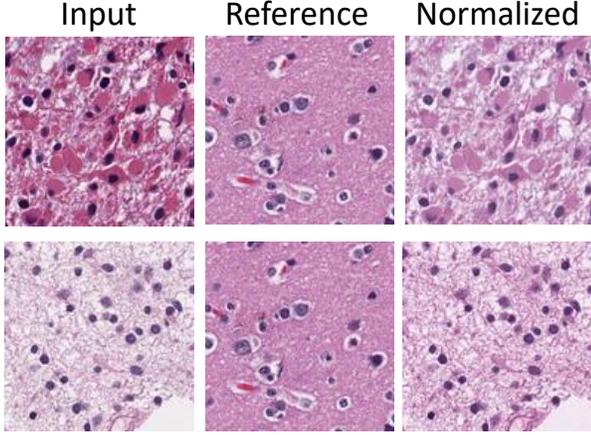


Figure 3: Color Normalization Visualization

We propose employing the same strategies to convert the input images to a same color domain in order to derive the diffusion model focus on learning the morphological patterns and other vital pathology aspects such as cell shape, density, and distribution rather than stain differences.

For color normalization, we used the structure-preserving color normalization scheme introduced by Vahadane et al.[39] that transfers source images to the target domain while preserving their own stain concentration. A Complete theoretical description of color normalization module is available in section 1.1 of supplementary material.

Figure 3 visualizes the performance of the color normalization method on three extracted patches of histopathology images for a same reference.

3.3. Diffusion Probabilistic Model

The diffusion model can be summarized in two main processes: forward diffusion and parameterized reverse diffusion. Figure 4 illustrates the two directions of diffusion probabilistic models. The model in the former process gradually generates noisier samples from real data using Gaussian noise kernel. The later process makes the model able to iteratively retrieve data from noise which can be employed to produce synthetic data from random noise.

3.3.1 Forward Diffusion

Let x_{0,g_m} be the real input data from the m -th genotype g_m and x_{t,g_m} be the noisy images for g_m produced at time $t = 1, 2, \dots, T$. Latent x_{t,g_m} can be derived directly from x_{0,g_m} as following [16]:

$$x_{t,g_m} = \sqrt{\alpha_t}x_{0,g_m} + \sqrt{1 - \alpha_t}\epsilon, \quad (5)$$

$0 < \beta_1, \beta_2, \dots, \beta_T < 1$ are fixed noise scales for each time step t and $\alpha_t := \prod_{s=1}^t (1 - \beta_s)$. Also, the distribution of these is as $\epsilon \sim \mathcal{N}(0, I)$. (Details are in section 1.2 of supplementary material)

3.3.2 Parametrized Reverse Diffusion

In order to generate a random sample in the reverse process, the latent x_{t,g_m} needs to be roughly an isotropic Gaussian distribution. The diffusion probabilistic model can be viewed similar to variational auto-encoders (VAE)[7], where the reverse process p_θ is learned by a neural network (section 3.3.5) and is equivalent to the decoder network in VAE. Contrary to VAE, the encoder in the diffusion model is a fixed forward diffusion process.

In the reverse process, our neural network ϵ_θ with parameters of θ learns to denoise the given x_{t,g_m} and output the x_{t-1,g_m} . With iterative subtraction of the noise predicted by the neural network (ϵ_θ), and starting with x_{t-1,g_m} which have standard Gaussian distribution, x_{T,g_m} can be written as[16]:

$$x_{t-1,g_m} = \mathbb{C}_1(x_{t,g_m} - \mathbb{C}_2\epsilon_\theta(x_{t,g_m}, t, g_m)) + \sigma_t z, \quad (6)$$

where:

$$\mathbb{C}_1 = (\sqrt{1 - \beta_t})^{-1}, \mathbb{C}_2 = \beta_t(\sqrt{1 - \alpha_t})^{-1}, \beta_t = \alpha_t^2 \quad (7)$$

Similarly, x_{0,g_m} which is the generated image at the end of iterations, can be written as:

$$x_{0,g_m} = \frac{1}{\sqrt{1 - \beta_1}} \left(x_{1,g_m} - \frac{\beta_1}{\sqrt{1 - \alpha_1}} \epsilon_\theta(x_{1,g_m}, 1, g_m) \right) + \sigma_1 z, \quad (8)$$

3.3.3 Training Loss

The final objective for training the utilized Diffusion probabilistic model is a combination of score matching losses [40] that can be summarized as the following:

$$Loss = L_{simple} + cL_{vlb}, \quad (9)$$

where:

$$L_{simple} = \sum_t \lambda_t L_t, \quad L_{vlb} = \sum_t L_t, \quad (10)$$

L_t is a score matching loss for the time step t which looks at the difference between the two Gaussian distributions. It can be written as:

$$L_t = D_{KL}(q(x_{t-1,g_m}|x_{t,g_m}, x_{0,g_m})|p_\theta(x_{t-1,g_m}|x_{t,g_m})) = \mathbb{E}_{x_{0,g_m}, \epsilon} \left[\frac{\beta_t}{(1 - \beta_t)(1 - \alpha_t)} \|\epsilon - \epsilon_\theta(x_{t,g_m}, t)\|^2 \right], \quad (11)$$

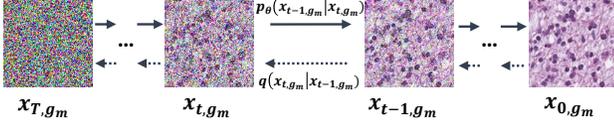


Figure 4: The illustration of the forward(0 to T) and reverse(T to 0) diffusion process

L_{simple} was initially proposed by Ho et al. [16] and use the following wights:

$$\lambda_t = \frac{(1 - \beta_t)(1 - \alpha_t)}{\beta_t}. \quad (12)$$

Considering the λ_t values, L_{simple} refers to a mean-squared error(MSE) loss defined on the difference of the actual and estimated noise, but Nicole et al. [32] added the second term to the loss function to learn the σ_t and showed that a small value for c can significantly improve the model’s capacity.

3.3.4 Morphology Levels Prioritization

Signal-to-noise-ratio (SNR) of the noisy image at the time step t (x_{t, g_m}) based on Equation 5 is equivalent to the following:

$$SNR(t) = \frac{\alpha_t}{1 - \alpha_t}, \quad (13)$$

Given the diminishing nature of SNR(t), it is demonstrated that the model concentrates rough and coarse properties during the early phases of the reverse diffusion process (when SNR is lower). Then, in the middle steps, it focuses on the image’s perceptual components, while the latter stages (with the highest SNR) are dedicated to imperceptible minutiae [7]. Histology images are fairly sensitive that requires more accurate features. Similarly, our model should be focused on learning pathological and morphological markers that pathologists need to make a diagnosis at intermediate steps before performing minor denoising tasks at the end. For morphology prioritization, the λ_t weights can be utilized to devote heavier weights to the loss at earlier levels to emphasise perceptual contents and lower weights to the later levels. We observed empirically that perception prioritized weighting provided by Choi et al. [7] can result in generating higher detailed histopathology images:

$$\lambda'_t = \frac{\lambda_t}{(k + SNR(t))^\gamma}, \quad (14)$$

where k and γ are used to keep the λ'_t from extraordinarily increasing for very low SNR values and to control the concentration on clean-up details, respectively.

3.3.5 The Architecture

We chose the backbone neural network similar to the Unet based model improved by Dhariwal et al. [10], which is inspired from the Unet model introduced by Ho et al. [16] for diffusion models. This model contains attention at three various resolutions that allows the model to concentrate on tiny features related to cells(e.g., cell shape or small blood veins) or larger elements like how cell distribution, the texture of the stroma or the overlaying tissue. It also benefits from BIGGAN downsampling/upsampling residual blocks[5] to maintain the model free of artefacts like checker boxes or aliasing, which may not be a vital issue for typical images but can completely disrupt the subtle and accurate patterns that should exist in histopathological images. It also uses embedding layer to inject timestep to the neural network. The rest of the weights of the model are shared between all the time steps. Moreover, genotypes are given to the model with a separated embedding layer similar to timesteps.

4. Experimental Evaluation and Results

In this section, we assess the performance of the proposed approach for utilizing diffusion probabilistic models on generating synthetic histopathology images and compare it against one of the closest works.

4.1. Data

We utilized a dataset of 344 whole slide images (WSIs) of low grade gliomas representative of its three major genomic subtypes from the Cancer Genome Atlas (TCGA) archive [14]. The dataset includes 297 cases with IDH mutations and 47 IDH Wild Type cases. The IDH Wild Type group has no IDH mutations and is labeled as IDHWT. Furthermore, the 297 IDH mutant slides are further divided into two groups: with no 1p19q chromosomal codeletion (173 slides) and with 1p19q codeletion (124 slides) labeled as IDHNC and IDHC, respectively. Each WSI is a large scale image with the size of $\sim 100K \times 100K$ pixels. Moreover, mutations associated with each patient were obtained from cbiportal (<https://www.cbiportal.org/>).

Each slide is pixel-wise annotated with an emphasis on the tumor-rich areas and attempted to avoid artefacts and empty spaces by a board-certified pathologist or a pathology resident under the supervision of a board-certified pathologist using our online annotation tool. These annotations will be made available to other researchers.

Annotated tumor areas from each slide were divided into small image tiles (referred to as patches) at specified objective magnification levels to improve computing performance. A maximum of 100 512×512 pixel patches from the tumor annotated regions were taken from each slide at original magnification of $40\times$ with a stride of 512 and

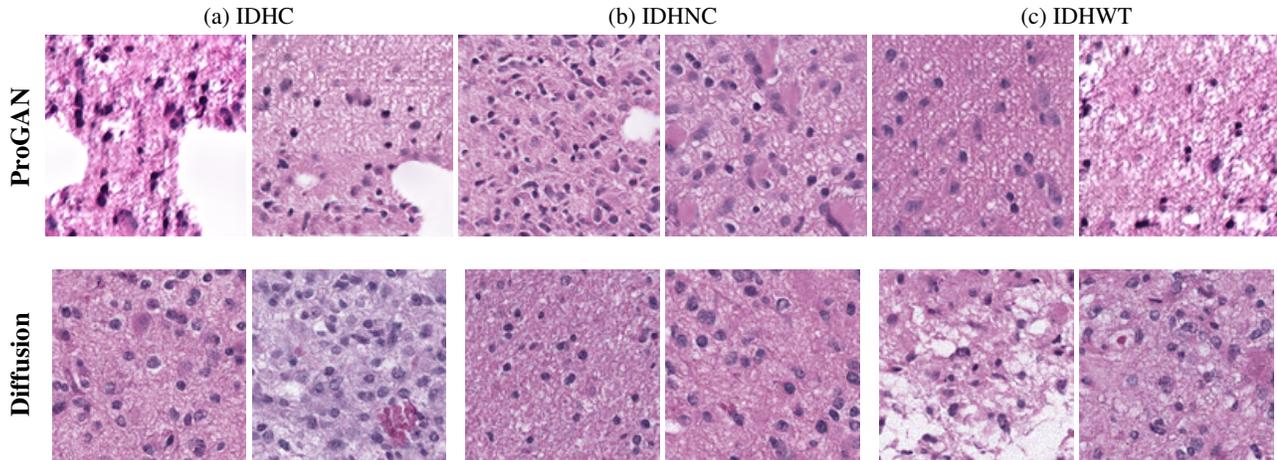


Figure 5: Selection of generated patches with diffusion and ProGAN models. First row represents synthesized images by ProGAN and second row refers images produced by diffusion model. Each dual columns together shows the generated samples representing one of the three different subtypes, namely IDHC, IDHNC, and IDHWT.

scaled to 128×128 patches, resulting in a final magnification of $10\times$. The pixel size at full-resolution was $\sim 0.25\mu\text{m}$ and down sampled to $\sim 1\mu\text{m}$. Finally, a total of 33,777 128×128 pixel patches (at $10\times$ magnification) were extracted from the WSIs and used to train various conditional diffusion probabilistic models. Table 2 and Figure 1 in supplementary material provide the breakdown of the extracted patches based on genomic subtypes and show examples of extracted patches.

4.2. Experiments

We evaluate the model’s performance in different unique scenarios to thoroughly examine the various objectives that the model should achieve. We also utilize several objective metrics to assess the quality of generated images based on each experiment’s specific requirements. Model implementation and training details are available at Table 1, section 1.2 in the supplementary material.

4.2.1 Experiment I

The objective of this experiment is to compare and contrast the quality of the synthesized images by our diffusion probabilistic model against a state-of-the-art study in which Levine et al. [23] utilized ProGAN [19]. In their study, the authors showed the superiority of histology images generated by ProGAN relative to other generative models such as variational autoencoder [11], enhanced super resolution GAN (ESRGAN) [42], and deep texture synthesis [12].

For a fair comparison, we utilized similarly normalized patches and due to the nature of our problem, we, inspired by [28], slightly modified ProGAN to generate histology images conditioned on genotypes. We also trained

both models on all the available extracted patches. We present samples of synthetic images generated by both models in Figure 5 (more images at Figure 2 in supplementary material). By zooming in this figure, unfavorable artefacts are clearly noticeable in the ProGAN generated images, especially in the top first and last images that cells are completely distorted and deformed. These samples demonstrate higher quality of the images synthesized by our model compared with those generated by ProGAN.

Next, we compared the two models by randomly generating 50,000 images by each model and calculating two sets of metrics:

1. Common Generative Evaluation Metrics: Three of the most widely used metrics for assessment of the generated images are: Inception Score (IS), Fréchet Inception Distance (FID), and sFID. We briefly discuss them in the following:

Inception Score (IS): We report Inception Score [36], which is defined as:

$$IS = \exp [E_{x \sim p_g} D_{KL}(p(y|x)||p(y))], \quad (15)$$

where $p(y)$ is marginal class probability and D_{KL} is the KL-divergence.

Fréchet inception distance (FID): This metric compares the distribution of generated images with the real images’ distribution in Inception-V3 latent space[20]. The more similar the synthetic images are to the input patches, the lower value that FID will have. The real and synthetic data are fed into the inception V3 model, and FID compares the

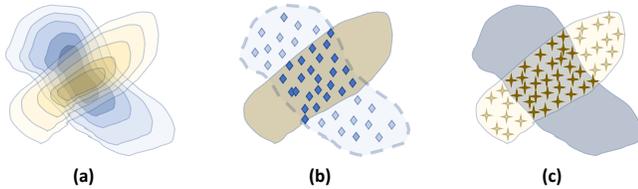


Figure 6: Real and synthetic data are depicted by blue and brown colors, respectively. (a) Presents the manifold for real and generated distributions. (b), and (c) Illustrate the Improved Recall and Precision, respectively.

mean and the standard deviation of the features extracted from pool_3 layer. The FID is given by:

$$FID(\mu_r, \Sigma_r, \mu_g, \Sigma_g) = \|\mu_r - \mu_g\|^2 + Tr(\Sigma_r + \Sigma_g - 2(\Sigma_r \Sigma_g)^{\frac{1}{2}}), \quad (16)$$

where μ_r and μ_g are the mean of the real and synthetic samples' embeddings. Similarly, Σ_r and Σ_g refer to their covariance.

sFID: This is a modified version of FID proposed by Szegedy et al. [38] that uses the initial channels from an intermediate layer to compare the means and standard deviations.

IS may not be a suitable statistic for generative models trained on datasets other than ImageNet, as noted by Barratt et al. [3]. As a result, evaluating the model using FID and sFID is critical. FID is less sensitive to spatial heterogeneity since it is calculated using features from one of the latest layers that compresses spatial information. However, sFID employs intermediate features, which can detect spatial similarity better than FID in some situations [30]. Reporting these together can measure the quality of the generated samples, which are summarized in Table 1. The results indicate that the proposed diffusion model outperforms the state-of-the-art across all these metrics. Also, lower values of both FID and sFID with extracted features from different layers make them sensitive to small changes and is able to detect mode coverage. This shows that unlike ProGAN the diffusion model is capable of producing perceptual features robustly.

| | ProGAN | Diffusion Model |
|-----------------|--------|-----------------|
| Inception Score | 1.67 | 2.08 |
| FID | 53.85 | 20.11 |
| sFID | 24.37 | 6.32 |

Table 1: Summary of the Inception Score, FID, and sFID for the first experiment

2. Improved Precision and Recall Metrics: Kynkäänniemi et al. [21] discussed that both the quality

and distribution coverage of the produced samples are essential for evaluating generative models. The authors proposed two metrics; namely "Improved Recall" and "Improved Precision" that can estimate both attributes by constructing non-parametric approximations of real and synthetic domain manifolds. We begin by estimating feature manifolds by computing distances to k-NN for each sample. Following that, "Improved Precision" refers to the percentage of produced samples inside the actual data manifold, while "Improved Recall" refers to the ratio of real samples located in the synthetic manifold.

These two concepts are depicted in Figure 6, and the results are summarized in Table 2. We can conclude that the proposed method produces better images than the state-of-the-art in terms of both diversity and fidelity. Also, it shows that our model is able to significantly differentiate morphological features of histology images.

| | ProGAN | Diffusion Model |
|--------------------|--------|-----------------|
| Improved Recall | 0.4816 | 0.8528 |
| Improved Precision | 0.0078 | 0.2573 |

Table 2: Summary of the Improved Recall and Precision for the first experiment

4.2.2 Experiment II

The purpose of this experiment is to compare the morphological properties of synthetic and actual images. We selected an equal number of real and synthetic images generated by our diffusion model and designed a pathologist survey consisting of the following two questions. The first question asks if the participants believe the image is real or synthetic, and the second question inquires about their confidence level (more details on the survey are available at Figure 3 and Figure 4, Section 1.5 in supplementary material). The images were displayed according to a random order. Two pathologists participated with varying levels of expertise participated in this survey: a board-certified pathologist (P1) and a pathology resident (P2). The summary of the results is given in Table 3, which shows that all participating experts could not distinguish the real from synthetic images generated by our diffusion model. For the majority of small percentage of synthetic images that experts were able to correctly identify, they indicated less confidence level. Our survey results show that our synthetic histopathology images look extremely similar to real examples, making them an excellent candidate for a variety of real-world applications.

We also utilized two sided Fisher-exact test to examine whether there is a statistically significant difference between each pathologist observations for the real and synthetic images (p-values are available in Table 4). The resulting p-values demonstrate there is no statistically signifi-

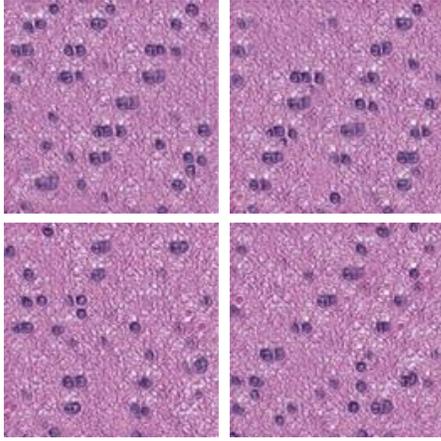


Figure 7: Examples of failed cases generated by ProGAN

cant difference between performance of the pathologists on identifying real versus synthetic images.

Table 3: Summary of results for Exp. II

| Conf. | Real High | Real Med. | Syn. Med. | Syn. High | Real All | Syn. All |
|---------|-----------|-----------|-----------|-----------|--------------|--------------|
| Real GT | 0.75 | 0.05 | 0.175 | 0.025 | 0.8 | 0.2 |
| Syn. GT | 0.775 | 0.05 | 0.125 | 0.05 | 0.825 | 0.175 |

(a) Summary of the results for P1

| Conf. | Real High | Real Med. | Syn. Med. | Syn. High | Real All | Syn. All |
|---------|-----------|-----------|-----------|-----------|--------------|--------------|
| Real GT | 0.225 | 0.2 | 0.25 | 0.325 | 0.425 | 0.575 |
| Syn. GT | 0.325 | 0.25 | 0.2 | 0.225 | 0.575 | 0.425 |

(b) Summary of the results for P2

| | P1 | P2 |
|------------------------|-----|---------|
| Fisher-exact's p-value | 1.0 | 0.26347 |

Table 4: Summary of the fisher test results

4.3. Visual Observation:

The bottom row in Figure 5 shows the images generated by our diffusion model. In these images, cell nuclei stained by purplish blue and extracellular matrix and cytoplasm stained by pink color related to H&E staining, which suggests that using the color normalization module is effective. Although medical inspection of generated images should be done by pathologists, there are some well known histology features in LGG images that can be identified even by an untrained person. The so-called "fried egg" appearance [9] of oligodendrogliomas is shown in the synthetic

images (such as the bottom first and second images in Figure 5, top second image or bottom fourth image at Figure 2 in the supplementary material), namely in the IDHC and IDHWT. Another characteristic of the oligodendrogliomas that can be seen in generated images by diffusion (such as the bottom second column of Figure 5) is branching small, chicken wire-like blood vessels [9], and this characteristic can also be found in the IDHC. This suggests that the diffusion model was able to learn specific known histopathological features. However, such specific features do not clearly exist in the images generated by ProGAN. In addition, the IDHWT are the most uncommon among the other two, and it appears that the lesser number of cases for the IDHWT subtype resulted in artefacts and lower image quality of ProGAN as compared to the diffusion, implying a mode collapse in ProGAN (Figure 7) due to a lack of enough data points in this subtype. However, the diffusion model could learn these rare class-specific features. Figure 7 shows samples of failed images by ProGAN in which each image is produced from random noise; however, they are fairly similar to each other.

5. Conclusion and Future Work

We proposed an end-to-end method based on diffusion probabilistic models to generate H&E stained histopathology images. To our knowledge, this is the first work that utilizes such models for histopathology image synthesis. Using multiple objective and subjective metrics, we compared the performance of our proposed approach to proGAN, that has shown remarkable performance in generating histopathology images. Results suggest that our proposed approach outperforms proGAN. Additionally, we conducted an empirical study where pathologists participated in a survey in which they were not able to distinguish the synthetic from real images. Taken together, the proposed method could facilitate the deployment of synthesized histology images for many real-life educational, privacy, and data augmentation applications.

In addition, the work in this paper can be extended by optimising the proposed model to reduce the sampling time of the diffusion probabilistic models, which is relatively longer than GANs due to multiple small diffusion steps. As an instance, this work can be expanded by drawing on the work of Xiao et al. [43], who use a multimodal conditioned discriminator that follows the diffusion model and can significantly reduce the number of diffusion steps.

6. Acknowledgment

This work is supported by the NSERC Discovery and CIHR Project grants.

References

- [1] Hani A Alturkistani, Faris M Tashkandi, and Zuhair M Mohammedsaleh. Histological stains: a literature review and case study. *Global journal of health science*, 8(3):72, 2016.
- [2] Tomer Amit, Eliya Nachmani, Tal Shaharabany, and Lior Wolf. Segdiff: Image segmentation with diffusion probabilistic models. *CoRR*, abs/2112.00390, 2021.
- [3] Shane Barratt and Rishi Sharma. A note on the inception score. *arXiv preprint arXiv:1801.01973*, 2018.
- [4] Jeffrey Boschman, Hossein Farahani, Amirali Darbandsari, Pouya Ahmadvand, Ashley Van Spankeren, David Farnell, Adrian B Levine, Julia R Naso, Andrew Churg, Steven JM Jones, et al. The utility of color normalization for ai-based diagnosis of hematoxylin and eosin-stained pathology images. *The Journal of Pathology*, 256(1):15–24, 2022.
- [5] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale GAN training for high fidelity natural image synthesis. In *International Conference on Learning Representations*, 2019.
- [6] Michele Ceccarelli, Floris P Barthel, Tathiane M Malta, Thais S Sabedot, Sofie R Salama, Bradley A Murray, Olena Morozova, Yulia Newton, Amie Radenbaugh, Stefano M Pagnotta, et al. Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma. *Cell*, 164(3):550–563, 2016.
- [7] Jooyoung Choi, Jungbeom Lee, Chaehun Shin, Sungwon Kim, Hyunwoo Kim, and Sungroh Yoon. Perception prioritized training of diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11472–11481, 2022.
- [8] Genevieve M Crane and Jerad M Gardner. Pathology image-sharing on social media: recommendations for protecting privacy while motivating education. *AMA journal of ethics*, 18(8):817–825, 2016.
- [9] Catherine Dumas-Duport, Pascale Varlet, Marie-Louise Tucker, Frederic Beuvon, Pascale Cervera, and Jean-Paul Chodkiewicz. Oligodendrogliomas. part i: Patterns of growth, histological diagnosis, clinical and imaging correlations: a study of 153 cases. *Journal of neuro-oncology*, 34(1):37–59, 1997.
- [10] Prafulla Dhariwal and Alex Nichol. Diffusion models beat gans on image synthesis. *CoRR*, abs/2105.05233, 2021.
- [11] Carl Doersch. Tutorial on variational autoencoders. *arXiv preprint arXiv:1606.05908*, 2016.
- [12] Leon Gatys, Alexander S Ecker, and Matthias Bethge. Texture synthesis using convolutional neural networks. *Advances in neural information processing systems*, 28, 2015.
- [13] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *arXiv*, 2014.
- [14] Robert L. Grossman, Allison P. Heath, Vincent Ferretti, Harold E. Varmus, Douglas R. Lowy, Warren A. Kibbe, and Louis M. Staudt. Toward a shared vision for cancer genomic data. *New England Journal of Medicine*, 375(12):1109–1112, 2016.
- [15] Feng Gu, Nikolay Burlutskiy, Mats Andersson, and Lena Kalland Wilén. Multi-resolution networks for semantic segmentation in whole slide images. In *Computational Pathology and Ophthalmic Medical Image Analysis*, pages 11–18. Springer, 2018.
- [16] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–6851, 2020.
- [17] Stephan W Jahn, Markus Plass, and Farid Moinfar. Digital pathology: advantages, limitations and emerging perspectives. *Journal of Clinical Medicine*, 9(11):3697, 2020.
- [18] Ansh Kapil, Armin Meier, Aleksandra Zuraw, Keith E Steele, Marlon C Rebelatto, Günter Schmidt, and Nicolas Brieu. Deep semi supervised generative learning for automated tumor proportion scoring on nsclc tissue needle biopsies. *Scientific reports*, 8(1):1–10, 2018.
- [19] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. *arXiv preprint arXiv:1710.10196*, 2017.
- [20] Tuomas Kynkäänniemi, Tero Karras, Miika Aittala, Timo Aila, and Jaakko Lehtinen. The role of imagenet classes in fréchet inception distance. *arXiv preprint arXiv:2203.06026*, 2022.
- [21] Tuomas Kynkäänniemi, Tero Karras, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Improved precision and recall metric for assessing generative models. *Advances in Neural Information Processing Systems*, 32, 2019.
- [22] Lan Lan, Lei You, Zeyang Zhang, Zhiwei Fan, Weiling Zhao, Nianyin Zeng, Yidong Chen, and Xiaobo Zhou. Generative adversarial networks and its applications in biomedical informatics. *Frontiers in public health*, 8:164, 2020.
- [23] Adrian B Levine, Jason Peng, David Farnell, Mitchell Nurse, Yiping Wang, Julia R Naso, Hezhen Ren, Hossein Farahani, Colin Chen, Derek Chiu, et al. Synthesis of diagnostic quality cancer pathology images by generative adversarial networks. *The Journal of pathology*, 252(2):178–188, 2020.
- [24] David N. Louis, Arie Perry, Guido Reifenberger, Andreas von Deimling, Dominique Figarella-Branger, Webster K. Cavenee, Hiroko Ohgaki, Otmar D. Wiestler, Paul Kleihues, and David W. Ellison. The 2016 world health organization classification of tumors of the central nervous system: a summary. *Acta Neuropathologica*, 131(6):803–820, 2016.
- [25] M. Y. Lu, R. J. Chen, D. Kong, J. Lipkova, R. Singh, D. Williamson, T. Y. Chen, and F. Mahmood. Federated learning for computational pathology on gigapixel whole slide images. *Medical Image Analysis*, 76:102298, 2022.
- [26] Troy Luhman and Eric Luhman. Diffusion models for handwriting generation. *ArXiv*, abs/2011.06704, 2020.
- [27] Puneet Mathur, Meghna P Ayyar, Rajiv Ratn Shah, and Shree G Sharma. Exploring classification of histological disease biomarkers from renal biopsy images. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 81–90. IEEE, 2019.
- [28] Takeru Miyato and Masanori Koyama. cgans with projection discriminator. *arXiv preprint arXiv:1802.05637*, 2018.
- [29] Pooya Mobadersany, Safoora Yousefi, Mohamed Amgad, David A. Gutman, Jill S. Barnholtz-Sloan, José E. Velázquez

- Vega, Daniel J. Brat, and Lee A. D. Cooper. Predicting cancer outcomes from histology and genomics using convolutional networks. *Proceedings of the National Academy of Sciences*, 115(11):E2970–E2979, 2018.
- [30] Charlie Nash, Jacob Menick, Sander Dieleman, and Peter W Battaglia. Generating images with sparse representations. *arXiv preprint arXiv:2103.03841*, 2021.
- [31] Cancer Genome Atlas Research Network. Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. *New England Journal of Medicine*, 372(26):2481–2498, 2015.
- [32] Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International Conference on Machine Learning*, pages 8162–8171. PMLR, 2021.
- [33] Vadim Popov, Ivan Vovk, Vladimir Gogoryan, Tasnima Sadekova, and Mikhail A. Kudinov. Grad-tts: A diffusion probabilistic model for text-to-speech. *CoRR*, abs/2105.06337, 2021.
- [34] J. Rajotte and R. Ng. Private data sharing between decentralized users through the privgan architecture. In *2020 IEEE 24th International Enterprise Distributed Object Computing Workshop*, pages 37–42, 2020.
- [35] Scott Ryall, Uri Tabori, and Cynthia Hawkins. Pediatric low-grade glioma in the era of molecular diagnostics. *Acta neuropathologica communications*, 8(1):1–22, 2020.
- [36] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training gans. *Advances in neural information processing systems*, 29, 2016.
- [37] Chetan L Srinidhi, Ozan Ciga, and Anne L Martel. Deep neural network models for computational histopathology: A survey. *Medical Image Analysis*, 67:101813, 2021.
- [38] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826, 2016.
- [39] Abhishek Vahadane, Tingying Peng, Amit Sethi, Shadi Albarqouni, Lichao Wang, Maximilian Baust, Katja Steiger, Anna Melissa Schlitter, Irene Esposito, and Nassir Navab. Structure-preserving color normalization and sparse stain separation for histological images. *IEEE Transactions on Medical Imaging*, 35(8):1962–1971, 2016.
- [40] Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural computation*, 23(7):1661–1674, 2011.
- [41] Sheng Wang, Jiawen Yao, Zheng Xu, and Junzhou Huang. Subtype cell detection with an accelerated deep convolution neural network. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 640–648. Springer, 2016.
- [42] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops*, pages 0–0, 2018.
- [43] Zhisheng Xiao, Karsten Kreis, and Arash Vahdat. Tackling the generative learning trilemma with denoising diffusion gans. In *International Conference on Learning Representations*, 2021.
- [44] Lipeng Xie, Jin Qi, Lili Pan, and Samad Wali. Integrating deep convolutional neural networks with marker-controlled watershed for overlapping nuclei segmentation in histopathology images. *Neurocomputing*, 376:166–179, 2020.
- [45] Yuan Xue, Jiarong Ye, Qianying Zhou, L Rodney Long, Sameer Antani, Zhiyun Xue, Carl Cornwell, Richard Zaino, Keith C Cheng, and Xiaolei Huang. Selective synthetic augmentation with histogan for improved histopathology image classification. *Medical image analysis*, 67:101816, 2021.
- [46] Qifan Zhou and Hujun Yin. A u-net based progressive gan for microscopic image augmentation. In *Annual Conference on Medical Image Understanding and Analysis*, pages 458–468. Springer, 2022.