Optimising Diffusion Models for Histopathology Image Synthesis

Victoria Porter[1]
vporter03@qub.ac.uk
Richard Gault[1]
richard.gault@qub.ac.uk
Stephanie Craig[2]
stephanie.craig@qub.ac.uk
Jacqueline James[2]
j.james@qub.ac.uk

[1] Centre for Intelligent Sustainable Computing, School of Electronics, Electrical Engineering and Computer Science Queen's University, Belfast

[2] Precision Medicine Centre, Patrick G. Johnston Centre for Cancer Research Queen's University, Belfast

Abstract

Oropharyngeal Squamous Cell Carcinoma (OPSCC) is a sub-type of head and neck cancer linked to human papillomavirus infection (HPV). HPV-positive OPSCC patients have an improved prognosis compared to HPV-negative OPSCC patients however, the reasoning for this is unknown. Visualising the clinical and molecular differences in HPV status would be highly interpretable and could aid our understanding of the impact these distinguishing features have on patient prognosis. A generative model trained to delineate features of HPV status provides both a synthetic visualisation of HPV-related OPSCC and a classification of HPV status. Conditional diffusion models (CDMs) have been shown to produce state-of-the-art (SOTA) quality and fidelity in the image synthesis domain. Furthermore, they can generate representative Haematoxylin and Eosin (H&E) stained histopathology images of cancerous tissue. This paper proposes two novel weighting schemes, one of which is designed to prioritise spatial features during training which enables the model to learn important pathological markers associated with HPV-related OPSCC tissue. Through experimental analysis of histological data, we demonstrate that our proposed approach improves the performance of CDMs and provides insightful, interpretable features that aid our understanding of HPV-related OPSCC.

1 Introduction

Head and neck cancer ranks as the 13th most common cancer globally when including cases of oropharyngeal cancer [1]. Squamous cell carcinoma is a sub-type of oropharyngeal cancer that is linked to high-risk variants of HPV. HPV is a DNA tumour virus that infects the skin and mucous membranes inside the body [4]. A study by Craig et al. [7] found that 37% of patients diagnosed with OPSCC in Northern Ireland also tested positive for HPV. Despite their histological similarities, HPV-positive OPSCC has substantial molecular and clinical differences compared to HPV-negative OPSCC [23]. Furthermore, HPV-positive OPSCC patients

have improved overall survival and significantly better progression-free survival rates compared to HPV-negative OPSCC patients [6, 7]; although the exact reasoning for this remains widely unknown. Further insight into the complex biology of the tumour microenvironment (TME) of HPV-related OPSCC is required to develop tailored treatment plans and improve overall survival and progression-free survival rates.

Generative models can create synthetic images that resemble the data on which they are trained. Synthetic images of tissue samples could be used to establish synthetic datasets as a proxy for real datasets. A generative model trained to learn the distinguishing perceptual and imperceptible features of HPV-positive and HPV-negative OPSCC may be applied to construct detailed synthetic images of HPV-related OPSCC. This could provide a deeper understanding of how HPV status affects the TME in OPSCC patients. Diffusion models are a class of generative models that have been shown to outperform Generative Adversarial Networks (GANs) [12] and Variational Autoencoders (VAEs) [18] in terms of sample quality [10]. Diffusion models learn a probability distribution over the input data by iteratively refining the distribution following the diffusion process. This allows the model to capture the underlying structure of the input data more effectively than GANs and VAEs, resulting in higher-quality samples [16]. Furthermore, diffusion models can be given context alongside the training data in the form of class labels or image embeddings which enables the model to generate samples of a particular type. These models are referred to as conditional diffusion models (CDMs), and have been shown to generate high-quality histopathology images of brain cancer [20]. Additionally, Dhariwal and Nichol [10] show that the sample quality of pre-trained CDMs can be improved further by exploiting a classifier trained on noisy images during the diffusion sampling process.

Choi et al. [5] demonstrate that diffusion models produce higher-quality samples when trained using their proposed perception prioritized (or P2) weighting scheme. Additionally, Moghadam et al. [20] applied the P2 weighting scheme to synthesise high-quality histopathology images. However, P2 was designed using the properties of natural images which significantly differ from the properties of histopathology images. In this paper, we propose alternative weighting schemes for CDMs with the objective of capturing relevant properties of histopathology images that may aid our exploration of the key morphological characteristics that relate HPV to OPSCC. Our contributions are as follows:

- We propose the exploitation of CDMs to generate synthetic images of HPV-related OPSCC tissue reflective of the pathological, molecular, and clinical differences that exist based on HPV status.
- We introduce two novel weighting schemes, bell-shaped and reverse-sigmoid, for the training objective function to encourage the CDM to learn perceptually discriminative features and imperceptible features respectively.
- We conduct a series of quantitative experiments using an HPV-related OPSCC dataset to evaluate the ability of the CDM to delineate HPV status in OPSCC tissue with and without classifier guidance.

2 Related Work

2.1 Diffusion Models for Image Synthesis

The diffusion process is comprised of two stages: forward diffusion and reverse diffusion. Forward diffusion involves iteratively adding noise to an input image through a sequence

of T time steps, where class labels or image embeddings can be provided as contextual auxiliary input. This context enables the model to capture class-specific properties from the data distribution which allows it to generate images that reflect these properties. Reverse diffusion employs a convolutional neural network trained to reverse the forward diffusion process over T timesteps to recover the original input. This process is applied to random noise during the diffusion sampling stage to produce synthetic images.

The breakthrough work on diffusion models by Ho et al. [16] utilised the framework proposed by Sohl-Dickstein et al. [28] to demonstrate that diffusion models can generate high-quality images at various resolutions. This requires the diffusion model to observe different levels of detail in the training images which enables it to learn perceptual and imperceptible features of the data and generate more realistic images. Dhariwal and Nichol [10] introduced architectural improvements to diffusion models based on the work of Ho et al. [16] and Song and Ermon [29] to demonstrate that diffusion models can outperform SOTA GANs in terms of image fidelity and diversity. Additionally, they trained a classifier on noisy images and used the resulting gradients to guide the diffusion sampling process towards an arbitrary class label which further improved the quality of generated samples. In theory, classifier guidance will encourage the model to replicate the structural heterogeneity and distinct morphological properties of OPSCC when conditioned on HPV status.

2.2 Diffusion Models in Histopathology

Diffusion models have been widely used for image generation tasks [11, 24, 27] since the improvements introduced by [10] and [16] however, the utilisation of such models within histopathology is limited. Moghadam et al. [20] employed a pipeline consisting of a colour normalisation module [31], a CDM, and the P2 weighting scheme [5] to synthesise Haematoxylin and Eosin (H&E) stained histopathology images of brain cancer. Consistent with the findings from [10], the results show that in terms of image quality, the diffusion model outperforms the SOTA Progressive Growing GAN [17] which has been shown to produce high-quality images of histopathology tissue [19]. The effect of classifier-free guidance on histopathology image synthesis was explored by [26] and demonstrated superior performance over [20]. Classifier-free guidance involves training both a conditional and an unconditional diffusion model and combining the resulting score estimates during the diffusion sampling process to achieve performance similar to that of classifier guidance [10] without the computational overhead of training a bespoke classifier. This work explores the effectiveness of classifier guidance on the synthesis of histopathology images and evaluates its impact relative to classifier-free guidance and unguided diffusion.

Assessing the quality of synthetic images is a critical aspect of evaluating the efficacy of generative models for histopathology image synthesis. In this work, the quality of randomly generated histopathology images will be assessed by HPVNet [8], a classifier designed to determine HPV status from H&E stained tissue of OPSCC. HPVNet will be used to provide an indicative measure of each model's ability to reproduce relevant biological markers associated with HPV status however, it is critical to emphasise that HPVNet cannot replace the expertise of a pathologist.

2.3 Redesigning the Training Objective Function

The diffusion process involves gradually corrupting an input image x_0 with pre-defined noise scales, $0 < \beta_1, \beta_2, ..., \beta_T < 1$, which correspond to a timestep t. The training objective of a

diffusion model can be expressed as the sum of denoising score matching losses [32] which measures how well the model can estimate the noise-free data distribution from the noisy data. The objective is to minimise these loss functions to improve the performance of the model. Ho et al. [16] proposed a simplified objective which was shown to improve sample quality and is defined as follows:

$$L_{simple} = \Sigma_t \lambda_t L_t, \tag{1}$$

where L_t is the denoising score matching loss for a timestep t and represents the distance between two Gaussian distributions, and λ_t refers to the weighting scheme which has uniform weights. λ_t is defined as follows:

$$\lambda_t = (1-\beta_t)(1-\alpha_t)/\beta_t, \tag{2}$$

where $\alpha_t := \prod_{s=1}^{t}(1-\beta_s)$. We refer to λ_t as the **default** weighting scheme herein. The training objective used for the CDMs in this work follows the hybrid objective, L_{hybrid}, proposed by Nichol et al. [22] which inherits the uniform weights from (2) and extends (1) by adding a second term, λL_{vlb}. This term represents the sum of denoising score matching losses across all timesteps and λ is assigned the value 0.001 to maintain L_{simple} as the most influential component. This configuration reduces the number of steps required during sampling without sacrificing sample quality compared to the objective defined by [16]. Formally, L_{hybrid} is defined as follows:

$$L_{hybrid} = L_{simple} + \lambda L_{vlb}, \tag{3}$$

Choi et al. [5] found that diffusion models learn features according to the signal-to-noise ratio (SNR) in three noise level stages which they define as coarse, content, and clean-up. They modified L_{simple} to force the model to focus more on the coarse and content stages to learn salient features and less on the clean-up stage where the model learns imperceptible features. Their proposed weighting scheme, λ'_t, is defined as:

$$\lambda'_t = \lambda_t/(k+SNR(t)^{\gamma}), \tag{4}$$

where the SNR of a corrupted image x_t is calculated by $SNR(t) = \alpha_t/(1-\alpha_t)$. We refer to λ'_t as the **P2** weighting scheme herein. Despite being optimised for the synthesis of natural images, the P2 weighting scheme has been shown to produce realistic histopathology images in [20] and [26]. In this work, we investigate the effect of prioritising alternative noise level stages during training on the synthesis of histopathology images. The aim is to establish a training objective that is optimised to capture distinguishing perceptual and imperceptible features of HPV-positive and HPV-negative OPSCC. A model that can highlight these distinguishing features would provide highly interpretable insights into the TME of HPV-related OPSCC. Such insights may aid pathologists' understanding of the effect of HPV status on OPSCC and support the development of tailored treatment plans.

3 Method

3.1 Conditional Diffusion Models

Dhariwal and Nichol[10] show that conditional diffusion improves sample quality compared to unconditional diffusion on the ImageNet dataset[9]. This is likely because ImageNet contains enough samples per class to enable the model to learn differentiating features between

classes, whereas unconditional models learn general features using the entire data distribution. Consistent with existing literature involving the synthesis of histopathology images [2, 13, 20, 26, 34], this work applies CDMs following the architecture provided by [10]. The training process is illustrated in Figure 1.

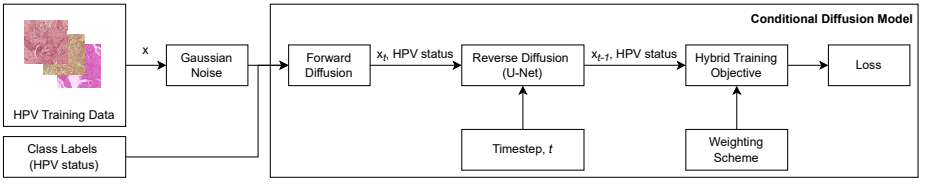

Figure 1: **Training pipeline**. A high-level overview of the training process for generating images of OPSCC tissue conditioned on HPV status.

3.2 Proposed Weighting Schemes

Two novel weighting schemes, bell-shaped and reversed-P2, are proposed using the three noise level stages described by Choi et al. [5]. The bell-shaped weighting scheme assigns higher weights to the content stage and lower weights to the course and clean-up stages. The primary objective of this weighting scheme is to encourage the model to prioritise perceptually discriminative content. In terms of HPV status, this may refer to pathological markers including p16 expression which indicates the presence of HPV infection [33]. Additionally, this weighting scheme places less emphasis on the course and clean-up stages. This prevents the model from focusing on course features such as colour and structure in addition to subtle pixel differences captured during clean-up. The bell-shaped weighting scheme is constructed as follows:

$$\text{Bell-Shaped}_t = \frac{1}{2} * (-cos(2\pi * t/T) + 1), \tag{5}$$

The reversed-P2 weighting scheme is the reverse arrangement of the P2 weighting scheme such that the highest weights are assigned to the content stage followed by the clean-up stage, with the course stage having the lowest assigned weights. The objective of this weighting scheme is to encourage the model to prioritise replicating perceptually discriminative features supplemented by visually indistinguishable details. The imperceptible details learned during the clean-up stage have little to no effect on the synthesis of high-fidelity natural images [5] however, such details within OPSCC tissue that are not easily identifiable by the untrained eye may be indicative of HPV status. The reversed-P2 weighting scheme is constructed as follows:

$$\text{Reversed-P2}_t = \lambda_t / (k + SNR(T + 1 - t))^{\gamma}), \tag{6}$$

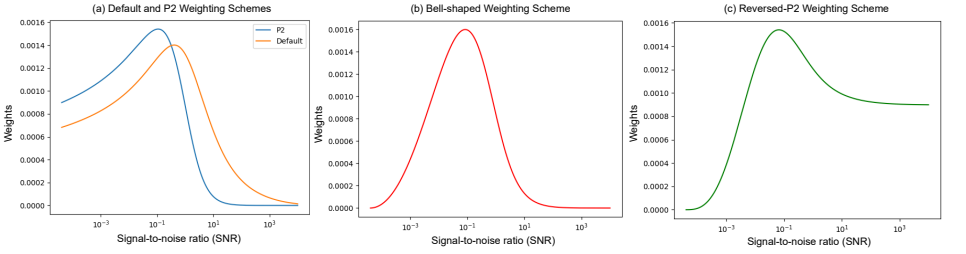

Figure 2: **Weighting Schemes**. Weights of the (a) Default and P2, (b) Bell-shaped, and (c) Reversed-P2 weighting schemes with a linear noise schedule.

3.3 Classifier Guidance

A noised-image classifier is specifically designed to classify images that have been intentionally corrupted with the addition of noise. The primary objective of these classifiers is to maintain robustness against the presence of noise in input images while still accurately classifying them into their respective classes. During the diffusion sampling process, the resultant gradients from the classifier are utilised in combination with the mean and covariance of image features provided by the CDM to produce a perturbed Gaussian noise distribution for each timestep t as illustrated in Figure 3. Dhariwal and Nichol [10] demonstrate that applying classifier guidance during the diffusion sampling process of CDMs significantly improves sample quality. In practice, individual datasets require the training of a bespoke classifier and Ho et al. [15] show that guidance by a classifier is limited in its effectiveness since the majority of the gradient information is irrelevant to predicting a class label. This work evaluates the impact of classifier guidance in the domain of histopathology image synthesis by assessing its effect on assisting the model in generating tissue samples of HPV-related OPSCC that are representative of intra-class and inter-class variations.

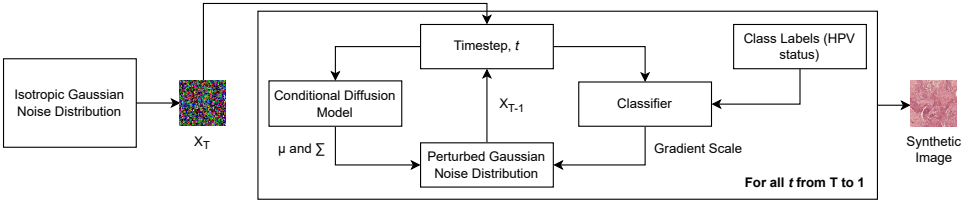

Figure 3: **Sampling pipeline**. A high-level overview of the sampling process with classifier guidance for generating images of OPSCC tissue conditioned on HPV status.

4 Experiment

4.1 Data

We utilised an HPV dataset containing 15,228 images of H&E-stained HPV-positive (56 patients, 9438 tiles) and HPV-negative (62 patients, 5790 tiles) OPSCC tissue extracted into 263x263 tiles from whole slide images. These images were resized and centre-cropped to a

resolution of 256x256 as part of the data pre-processing performed by the diffusion model. The default, P2, bell-shaped, and reversed-P2 weighting schemes were individually applied to train 4 CDMs on the HPV dataset for 250,000 steps. To evaluate the CDMs, a reference batch containing pre-computed statistics was created using the entire HPV dataset according to the procedure established in [10]. Additionally, an unlabelled and unseen HPV dataset of 9,468 HPV-positive (15 patients, 4,578 tiles) and HPV-negative (15 patients, 4,890 tiles) images was used to compute Precision and Recall. The CDMs were used to generate 50,000 synthetic samples each which were compared to the reference batch and the unseen HPV dataset to compute the evaluation metrics discussed in Section 4.2.

4.2 Evaluation Metrics

The primary evaluation statistics used for the CDMs include the Inception Score (IS)[25], Fréchet Inception Distance (FID)[14] and Spatial FID (sFID) [30]. IS was shown to correlate well with human scoring of the realism of synthetic images [3] such that a high IS is reflective of diverse images with high perceptual quality. However, as IS is entwined with the ImageNet dataset it may not be reflective of measuring pathologically relevant morphology that is important to a pathologist. FID and sFID are used in conjunction with IS to provide a more quantitative perspective. FID compares the distribution of synthetic images with the distribution of ground truth images whilst sFID considers the spatial information of the generated images in addition to comparing the real and synthetic image distributions. Low FID and sFID scores indicate that the model generates images that are similar to the real images. In particular, a low sFID indicates that the generated images reflect spatial features such as textures, structure, and layout that are consistent with real images.

4.3 Quantitative Comparison of Weighting Schemes

The objective of this experiment is to evaluate the quality and diversity of images generated by each of the four CDMs without classifier guidance on the HPV dataset. Each model was used to randomly generate 25,000 HPV-positive and 25,000 HPV-negative synthetic images. These batches were evaluated against the reference batch and an additional unseen dataset containing HPV-related OPSCC images. The results are available in Table 1.

Table 1: Performance of the CDMs on the HPV dataset.

Model	Guidance	IS ↑	FID ↓	sFID ↓
CDM-Default	✗	2.48	155.07	27.77
CDM-P2	✗	**3.09**	152.40	24.09
CDM-Reversed-P2	✗	2.92	191.56	31.15
CDM-Bell-Shaped	✗	2.93	**116.36**	**18.75**

The choice of weighting scheme is an important consideration for improving the performance of diffusion models in different applications. The bell-shaped weighting scheme scores the lowest sFID and FID scores meaning that the resulting synthetic samples best reflect the structural properties, such as textures, structures, and layout, of the real HPV data compared to the other weighting schemes. The P2 weighting scheme scores the highest IS which indicates high perceptual quality by learning coarse features such as the global colour

scheme however, it results in relatively high FID and sFID scores meaning that it is unable to extract meaningful spatial features compared to the bell-shaped weighting scheme. This demonstrates that spatial features captured by the bell-shaped scheme are significant for generating representative histopathology images of HPV-related OPSCC. Placing higher weights in the course stage, as done by the default and P2 weighting schemes, distracts the model from learning key morphological properties and perceptually discriminative details associated with spatial features. The reversed-P2 weighting scheme produces samples with the lowest relative quality and diversity. This is consistent with the findings of [5] whereby the model cannot capture perceptually rich information when a greater emphasis is placed on the clean-up stage. Equally, placing little to no emphasis on the coarse stage, as done by the reversed-P2 weighting scheme may affect the quality and convergence of the clean-up stage where the imperceptible details are learned. This demonstrates that focusing on imperceptible details alone in the clean-up stage is not significantly important for the model to delineate HPV status. Examples of the synthetically generated data are included in Figure 4.

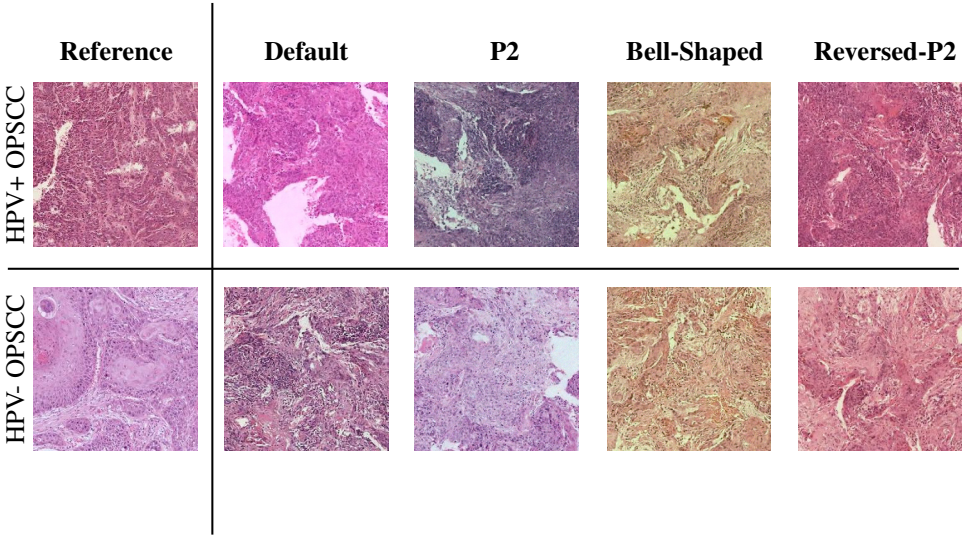

Figure 4: **CDM synthetic samples**: A random selection of synthetic images produced by the CDMs using each weighting scheme. Each weighting scheme displays an example of synthetic HPV+ OPSCC (top) and synthetic HPV- OPSCC (bottom).

4.4 The Effect of Classifier Guidance

The objective of this experiment is to investigate the effectiveness of incorporating classifier guidance during the sampling process of the CDMs described in Section 4.3 in terms of sample quality and diversity. These models are referred to as GCDMs for **G**uided CDMs. A 256x256 noised image classifier was pre-trained on the HPV dataset following the same format as [10]. The classifier was used alongside the CDMs during the diffusion sampling process to generate 50,000 images and corresponding labels for each weighting scheme. The results are available in Table 2 and examples of the synthetically generated data are included in 5.

Table 2: Performance of the classifier guided CDMs on the HPV dataset.

Model	Guidance	IS ↑	FID ↓	sFID ↓
CDM-Default	✓	2.50	152.35	27.34
CDM-P2	✓	**3.00**	146.37	23.11
CDM-Reversed-P2	✓	2.96	188.68	30.71
CDM-Bell-Shaped	✓	2.92	**115.90**	**18.45**

The results show that classifier guidance marginally improves the FID and sFID scores on the HPV dataset for all weighting schemes with the bell-shaped scheme producing the best results. The IS improves slightly for most of the weighting schemes except for P2 which shows a marginal decrease. This is expected since IS is not designed to measure the perceptual quality of pathological tissue. This raises the question as to whether training a separate classifier is worth the additional time and computation required. Furthermore, the addition of a classifier adds an extra layer of complexity to the model, which causes further delays in diffusion sampling time. The effectiveness of classifier guidance depends heavily on the quality and accuracy of the classifier used. If the classifier is not properly trained or is not well-suited to the target distribution, the results may be sub-optimal. Moreover, classifier guidance may not be effective for all types of data, especially if the target distribution is too complex or poorly defined.

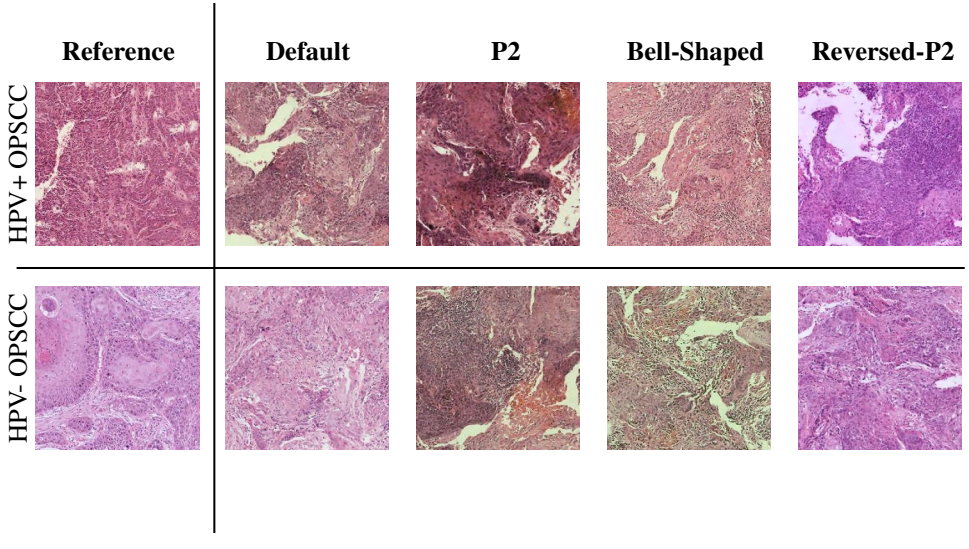

Figure 5: **GCDM synthetic samples**: A random selection of synthetic images produced by the GCDMs using each weighting scheme. Each weighting scheme displays an example of synthetic HPV+ OPSCC (top) and synthetic HPV- OPSCC (bottom).

4.5 Analysis of Feature Representation in Synthetic Samples

This experiment provides a quantitative analysis of the models' ability to represent relevant features of HPV-related OPSCC. HPVNet [8] is a classifier trained to learn distinguishing

features of HPV+ and HPV- tissue samples as part of a classification task. HPVNet is used in this experiment to assess the models' ability to generate images that are sufficiently representative of known HPV-related features to enable classification. To measure this, 25,000 HPV-positive and 25,000 HPV-negative samples were randomly generated for each of the models discussed in Sections 4.3 and 4.4. These samples were processed by HPVNet to produce the scores shown in Table 3.

Table 3: Classification of synthetic samples generated by each model using HPVNet.

Model	Guidance	Accuracy ↑	F1 ↑	Precision ↑	Recall ↑
CDM-Default	✗	0.6606	0.6421	0.6798	0.6083
	✓	0.6850	0.6651	**0.7097**	0.6258
CDM-P2	✗	0.6665	0.7005	0.6338	0.7830
	✓	0.6983	0.7253	0.6662	0.7960
CDM-Reversed-P2	✗	0.6436	0.6776	0.6198	0.7472
	✓	0.6604	0.6917	0.6355	0.7589
CDM-Bell-Shaped	✗	0.6904	0.7229	0.6561	0.8049
	✓	**0.7102**	**0.7415**	0.6698	**0.8304**

Consistent with the results in Sections 4.3 and 4.4, the bell-shaped weighting scheme with classifier guidance produces samples that consistently exhibit features that are representative of the inter-class differences associated with HPV status. The CDM-P2 models trained using the P2 weighting scheme with classifier guidance produced results that closely align with the unguided bell-shaped CDM indicating that even without classifier guidance the bell-shaped weighting scheme performs considerably well. Interestingly, the metrics suggest that the majority of the generated images contain features indicative of HPV status in OPSCC tissue for all models despite considerable colour variation in the HPV dataset. One possible explanation for this is that the HPVNet classifier was trained using multiple colour-based augmentations making it robust to the colour variation exhibited in the generated samples. Whilst HPVNet is a useful tool for evaluating the feature relevance in the generated images, it is not a replacement for a pathologist's expertise. A combination of both approaches is required to provide a comprehensive and reliable assessment.

5 Conclusion

We present two novel weighting schemes for training CDMs on histological data and demonstrate that insightful and interpretable features are reflected in the synthetic images. In particular, the bell-shaped weighting scheme enables the models to generate representative samples of HPV-related OPSCC tissue compared to existing methods [5]. In this work we apply HPVNet [8] as an independent assessment tool for the synthetically generated images however, future work should employ guidance from a pathologist for cross-validation. Furthermore, the results from HPVNet indicate that relevant pathological features are captured when using the proposed weighting schemes. Future work should explore these features with respect to the SNR to provide context on why the classification improves when prioritising the proposed noise levels. Finally, future work should address the colour inconsistencies associated with differing digital pathology scanners and staining protocols. This may be addressed using colour normalisation such as [31] or [21].

Acknowledgments

We would like to thank Prof. Iain Styles for providing valuable feedback, and Dr. Kris McCombe for extracting the tiles that comprise the HPV dataset.

References

[1] *Global oral health status report: towards universal health coverage for oral health by 2030*, chapter 2.4.1. World Health Organization, Geneva, 2022. Licence: CC BY-NC-SA 3.0 IGO.

[2] Marco Aversa, Gabriel Nobis, Miriam Hägele, Kai Standvoss, Mihaela Chirica, Roderick Murray-Smith, Ahmed Alaa, Lukas Ruff, Daniela Ivanova, Wojciech Samek, Frederick Klauschen, Bruno Sanguinetti, and Luis Oala. Diffinfinite: Large mask-image synthesis via parallel random patch diffusion in histopathology. In *Thirty-seventh Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2023.

[3] Shane T. Barratt and Rishi Sharma. A note on the inception score. In *Proceedings of the ICML 2018 Workshop on Theoretical Foundations and Applications of Deep Generative Models*, 2018.

[4] Kari P. Braaten and Marc R. Laufer. Human papillomavirus (hpv), hpv-related disease, and the hpv vaccine. *Reviews in Obstetrics and Gynecology*, 1(1):2–10, 2008.

[5] 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 (CVPR)*, pages 11472–11481, 2022.

[6] Anthony R. Cillo, Cornelius H. L. Kürten, Tracy Tabib, Zengbiao Qi, Sayali Onkar, Ting Wang, Angen Liu, Umamaheswar Duvvuri, Seungwon Kim, Ryan J. Soose, Steffi Oesterreich, Wei Chen, Robert Lafyatis, Tullia C. Bruno, Robert L. Ferris, and Dario A. A. Vignali1. Immune landscape of viral- and carcinogen-driven head and neck cancer. *Immunity*, 52(1):183–199, 2020.

[7] Stephanie G. Craig, Lesley A. Anderson, Andrew G. Schache amd Michael Moran, Laura Graham, Keith Currie, Keith Rooney, Max Robinson, Navdeep S. Upile, Rachel Brooker, Mina Mesri, Victoria Bingham, Stephen McQuaid, Terry Jones, Dennis J. McCance, Manuel Salto-Tellez, Simon S. McDade, and Jacqueline A. James. Recommendations for determining hpv status in patients with oropharyngeal cancers under tnm8 guidelines: a two-tier approach. *British Journal of Cancer*, 120(8):827–833, 2019.

[8] Stephanie G. Craig, Richard Gault, Kris McCombe, Andrew Moyes, Youcheng Sun, Tao Wang, Andrew G. Schache, Terry M. Jones, Janet M. Risk, Philip Gunning, Gaborieau, Valerie, Paul Brennan, Behnoush Abedi-Ardekani, and Jacqueline James. A deep learning alternative to regional molecular testing for hpv status. AACR-AHNS Head and Neck Cancer Conference: Innovating through Basic, Clinical, and Translational Research,

2023. URL https://www.aacr.org/professionals/meetings/previous-aacr-meetings/previous-aacr-meetings-2023/.

[9] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, 2009.

[10] Prafulla Dhariwal and Alexander Quinn Nichol. Diffusion models beat GANs on image synthesis. In *Advances in Neural Information Processing Systems*, volume 34, page 8780–8794, 2021.

[11] Tim Dockhorn, Arash Vahdat, and Karsten Kreis. Score-based generative modeling with critically-damped langevin diffusion. In *International Conference on Learning Representations*, 2022.

[12] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. In *Advances in Neural Information Processing Systems*, volume 27, page 2672–2680, 2014.

[13] Alexandros Graikos, Srikar Yellapragada, Minh-Quan Le, Saarthak Kapse, Prateek Prasanna, Joel Saltz, and Dimitris Samaras. Learned representation-guided diffusion models for large-image generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8532–8542, 2024.

[14] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Advances in Neural Information Processing Systems*, volume 30, 2017.

[15] Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. In *NeurIPS 2021 Workshop on Deep Generative Models and Downstream Applications*, 2021.

[16] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33:6840–851, 2020.

[17] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of GANs for improved quality, stability, and variation. In *International Conference on Learning Representations*, 2018.

[18] Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. In *International Conference on Learning Representations*, 2014.

[19] Adrian B. Levine, Jason Peng, David Farnell, Mitchell Nursey, Yiping Wang, Julia R. Naso, Hezhen Ren, Hossein Farahani, Colin Chen, Derek Chiu, Aline Talhouk, Brandon Sheffield, Maziar Riazy, Philip P. Ip, Carlos Parra-Herran, Anne Mills, Naveena Singh, Basile Tessier-Cloutier, Taylor Salisbury, Jonathan Lee, Tim Salcudean, Steven Jm. Jones, David G. Huntsman, C Blake Gilks, Stephen Yip, and Ali Bashashati. Synthesis of diagnostic quality cancer pathology images by generative adversarial networks. *The Journal of Pathology*, 252, 2020.

[20] Puria Azadi Moghadam, Sanne Van Dalen, Karina C. Martin, Jochen Lennerz, Stephen Yip, Hossein Farahani, and Ali Bashashati. A morphology focused diffusion probabilistic model for synthesis of histopathology images. In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 1999–2008, 2023.

[21] Andrew Moyes, Richard Gault, Kun Zhang, Ji Ming, Danny Crookes, and Jing Wang. Multi-channel auto-encoders for learning domain invariant representations enabling superior classification of histopathology images. *Medical Image Analysis*, 83:1361–8415, 2023.

[22] Alexander Q. Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139, pages 8162–8171, 2021.

[23] Steven F. Powell, Lexi Vu, William C. Spanos, and Dohun Pyeon. The key differences between human papillomavirus-positive and -negative head and neck cancers: Biological and clinical implications. *Cancers (Basel)*, 13(20):5206, 2021.

[24] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10674–10685, 2022.

[25] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training gans. In *Advances in Neural Information Processing Systems*, volume 29, 2016.

[26] Aman Shrivastava and P. Thomas Fletcher. Nasdm: nuclei-aware semantic histopathology image generation using diffusion models. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 786–796, 2023.

[27] Abhishek Sinha, Jiaming Song, Chenlin Meng, and Stefano Ermon. D2c: Diffusion-decoding models for few-shot conditional generation. In *Advances in Neural Information Processing Systems*, volume 34, pages 12533–12548, 2021.

[28] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *Proceedings of the 32nd International Conference on Machine Learning*, volume 37, pages 2256–2265, 2015.

[29] Yang Song and Stefano Ermon. Improved techniques for training score-based generative models. In *Advances in Neural Information Processing Systems*, volume 33, pages 12438–12448. Curran Associates, Inc., 2020.

[30] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2818–2826, 2016.

[31] Abhishek Vahadane, Tingying Peng, Amit Sethi, Shadi Albarqouni, Lichao Wang, Maximilian Baust, Katja Steiger, Anna M. 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.

[32] Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural Computation*, 23(7):1661–1674, 2011.

[33] Lee Fah Yap, Sook Ling Lai, Anthony Rhodes, Hans Prakash Sathasivam, Maizaton Atmadini Abdullah, Kin-Choo Pua, Pathmanathan Rajadurai, Phaik-Leng Cheah, Selvam Thavaraj, Max Robinson, and Ian C. Paterson. Clinico-pathological features of oropharyngeal squamous cell carcinomas in malaysia with reference to hpv infection. *Infectious Agents and Cancer*, 13, 2018.

[34] Srikar Yellapragada, Alexandros Graikos, Prateek Prasanna, Tahsin Kurc, Joel Saltz, and Dimitris Samaras. Pathldm: Text conditioned latent diffusion model for histopathology. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 5182–5191, 2024.