Controllable and Efficient Multi-Class Pathology Nuclei Data Augmentation using Text-Conditioned Diffusion Models

Hyun-Jic Oh and Won-Ki Jeong

College of Informatics, Department of Computer Science and Engineering,
Korea University, Seoul, South Korea
`wkjeong@korea.ac.kr`

Abstract. In the field of computational pathology, deep learning algorithms have made significant progress in tasks such as nuclei segmentation and classification. However, the potential of these advanced methods is limited by the lack of available labeled data. Although image synthesis via recent generative models has been actively explored to address this challenge, existing works have barely addressed label augmentation and are mostly limited to single-class and unconditional label generation. In this paper, we introduce a novel two-stage framework for multi-class nuclei data augmentation using text-conditional diffusion models. In the first stage, we innovate nuclei label synthesis by generating multi-class semantic labels and corresponding instance maps through a joint diffusion model conditioned by text prompts that specify the label structure information. In the second stage, we utilize a semantic and text-conditional latent diffusion model to efficiently generate high-quality pathology images that align with the generated nuclei label images. We demonstrate the effectiveness of our method on large and diverse pathology nuclei datasets, with evaluations including qualitative and quantitative analyses, as well as assessments of downstream tasks.

Keywords: Diffusion models · Data augmentation · Pathology Nuclei Segmentation and Classification.

1 Introduction

In digital pathology, nuclei segmentation and classification are essential tasks for accurate disease diagnosis and prognosis. These processes allow for the quantitative analysis of complex nucleus properties such as size, shape, and distribution [1]. Recent advances in deep learning-based methods have significantly improved these processes [9, 13]. However, the limited availability of labeled data remains a challenge, restricting the predictive capabilities of these algorithms. The time-consuming nature of pathology image acquisition and labeling further worsens this problem.

To overcome these challenges, data synthesis using deep generative models has been emerging as a promising solution. While generative adversarial networks

(GANs) [19, 26] have been widely employed for image generation [7, 25], recent studies show that diffusion models outperform GANs in natural images [4, 11, 27], making them a new standard method in many image synthesis tasks. These approaches aim to address the efficiency of data collection and enhance dataset diversity.

Although generative models have been successful in synthesizing natural images, we have observed several challenges in adopting this method for pathology image synthesis. For example, SDM [27] utilizes semantic labels as conditions for image synthesis, which is later extended to pathology nuclei image synthesis tasks [17, 23, 30]. Despite their ability to produce high-quality images, pixel-based diffusion models are less practical for data augmentation due to the time and computational cost required for training and inference procedures. Moreover, existing methods have mainly focused on image synthesis rather than realistic label augmentation; many nuclei data augmentation methods rely on a simple geometric alteration, such as random perturbation of nuclei positions [17], copy and pasting [5], or deformation [7], without consideration to the spatial context of the labels. Yu *et al.* [30] employed an unconditional diffusion model [11] that synthesizes nuclei labels with distance transform masks for both dense pixel semantic and instance label generation. However, their approach lacks scalability for multi-class data, and the standard Gaussian diffusion model is not optimal for multi-class semantic data synthesis. Instead, Hoogeboom *et al.* [12] suggested utilizing categorical distribution for semantic data synthesis, and Park *et al.* [18] introduced a joint Gaussian-categorical diffusion model that generates image-label pair synthesis. However, we observed that unconditional diffusion models tend to replicate frequently observed structures in the training data, resulting in biased class distribution, which is especially severe in pathology nuclei label synthesis.

In this paper, we propose a novel two-stage multi-class pathology nuclei data augmentation framework addressing the above issues. First, we develop a text-conditional joint diffusion model for spatially controllable multi-class nuclei label synthesis, inspired by the text-conditional generation model [21] that utilizes text-image alignment with target-related text prompts to generate fine-grained images. To achieve this, we utilize a pre-trained text encoder that is specifically designed for pathology images [29]. Second, we formulate a semantic and text conditional latent diffusion model (LDM) based on the pretrained LDMs (e.g., Zhang *et al.* [32]) and fine-tune it for pathology nuclei image synthesis, which significantly enhances performance and efficiency in the sampling process. We demonstrate the effectiveness of our method across various datasets stained with different modalities, conducting comprehensive evaluations that assess both the quality of synthesized data and the performance of downstream tasks. Our contributions can be summarized as follows:

– We introduce a novel two-stage framework for pathology nuclei data augmentation consisting of multi-class label synthesis and high-quality realistic image synthesis. We pioneer a multi-class dense pixel-by-pixel label generation strategy and tailor the pretrained LDMs to synthesize pathology
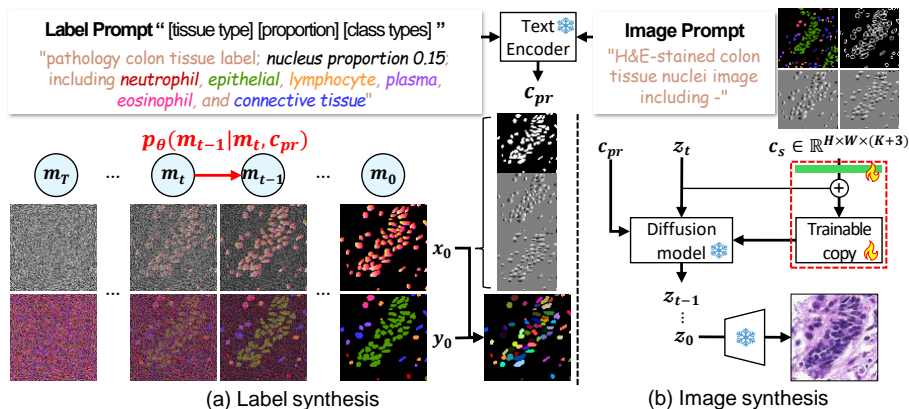
Fig. 1: Overview of the proposed two-stage data synthesis framework, consisting of label and image synthesis steps. The framework utilize spatial text condition c_{pr} for label synthesis (a). For image synthesis (b), we fine-tune the pretrained latent diffusion model with semantic condition c_s and text condition c_{pr}.

images corresponding to the generated nuclei label images, ensuring both image quality and computational efficiency.

– We propose text conditioning for detailed guidance of the label synthesis process to mitigate the inherent bias of unconditional models, which typically favor generating labels with nucleus proportions frequently observed in the training data. Our diffusion-based method implicitly learns the spatial layout of nuclei while text conditioning allows fine control of class distribution in the synthesized image, which effectively avoids data imbalance problems.

– We validate our methodology on large public datasets with a variety of staining modalities, and its effectiveness has been demonstrated through comprehensive qualitative and quantitative analysis, along with evaluation of downstream tasks.

2   Method

In this section, we introduce a novel two-stage data augmentation framework for multi-class pathology nuclei data using diffusion models. We provide background on diffusion models (Sec. 2.1), followed by detailed descriptions of our approach to text conditional label synthesis (Sec. 2.2) and LDM-based semantic and text conditional image synthesis (Sec. 2.3). The overview of the framework is illustrated in Fig. 1.

2.1   Background: Diffusion Models

Diffusion models [11] synthesize data from a learned data distribution by converting noise, operating on a forward and backward diffusion processes. In the

forward process $q(\mathbf{x}_t|\mathbf{x}_{t-1})$, noise ϵ_t is incrementally added to data $\mathbf{x}_0$ over timesteps $t \in \{1, ..., T\}$ according to a predefined noise schedule β_t, resulting in a pure noise $\mathbf{x}_T$ that aligns with a specific distribution. Then, the model ϵ_θ learns to predict the noise by the backward process $p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)$ to reconstruct $\mathbf{x}_0$. Standard diffusion models generally employ a Gaussian noise for ordinal data synthesis such as images. The diffusion process is then defined as:

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

$$p_\theta(\mathbf{x}_{t-1}|x_t) := \mathcal{N}(\mathbf{x}_{t-1}; \mu_\theta(\mathbf{x}_t), \sigma_t^2\mathbf{I}). \tag{2}$$

The training objective is then defined as:

$$\mathbb{E}_{x,\epsilon\sim\mathcal{N}(0,1),t}\left[\|\epsilon - \epsilon_\theta(\mathbf{x}_t, t)\|_2^2\right]. \tag{3}$$

For synthesizing categorical data $\mathbf{y}$ such as semantic labels, a categorical diffusion model is employed, using a categorical distribution $\mathcal{C}$ as defined by Hoogeboom *et al.* [12]:

$$q(\mathbf{y}_t|\mathbf{y}_{t-1}) := \mathcal{C}(\mathbf{y}_t; (1 - \beta_t)\mathbf{y}_{t-1} + \beta_t/K), \tag{4}$$

$$p_\theta(\mathbf{y}_{t-1}|\mathbf{y}_t) := \mathcal{C}(\mathbf{y}_{t-1}; \Theta_\theta(\mathbf{y}_t)), \tag{5}$$

where K represents the number of categories and Θ is the probability mass function of $\mathcal{C}$.

For pair-wise synthesis of ordinal and categorical data, a Gaussian-categorical diffusion model [18] formulates a joint distribution of Gaussian and categorical distributions $\mathcal{NC}$ as:

$$p_\theta(\mathbf{m}_{t-1}|\mathbf{m}_t) := \mathcal{NC}(\mathbf{m}_{t-1}; \mu_\theta(\mathbf{m}_t), \Sigma_\theta(\mathbf{m}_t), \Theta_\theta(\mathbf{m}_t)), \tag{6}$$

where $\mathbf{m}$ indicates the data pair $(\mathbf{x}, \mathbf{y})$.

2.2   Text Conditional Multi-class Label Synthesis

Synthetic nuclei label requires an corresponding instance label to support advanced nuclei analysis algorithms. Considering the difficulty of nucleus instance separation because of clustered nucleus, we formulate a text conditional joint diffusion model that simultaneously generates semantic label with structure map for nuclei separation. First, we define the semantic label as one-hot shaped vector $y \in \mathbb{R}^{H \times W \times K}$ with categorical diffusion process, where K indicates the number of classes. Additionally, we define a 3-channel structure map $x \in \mathbb{R}^{H \times W \times 3}$ as the concatenation of a binarized semantic label with x- and y-directional distance transform maps inspired by [30] with Gaussian diffusion process (see x_0 in Fig. 1a). We define a data pair $m_0 = (x_0, y_0) \in \mathbb{R}^{H \times W \times 6}$ by concatenating x_0 and y_0 along channel dimension, while y_0 is embedded into 3-channel via learnable parameters. Then, the joint diffusion model is trained to generate a data pair m_0 as Eq. 6.
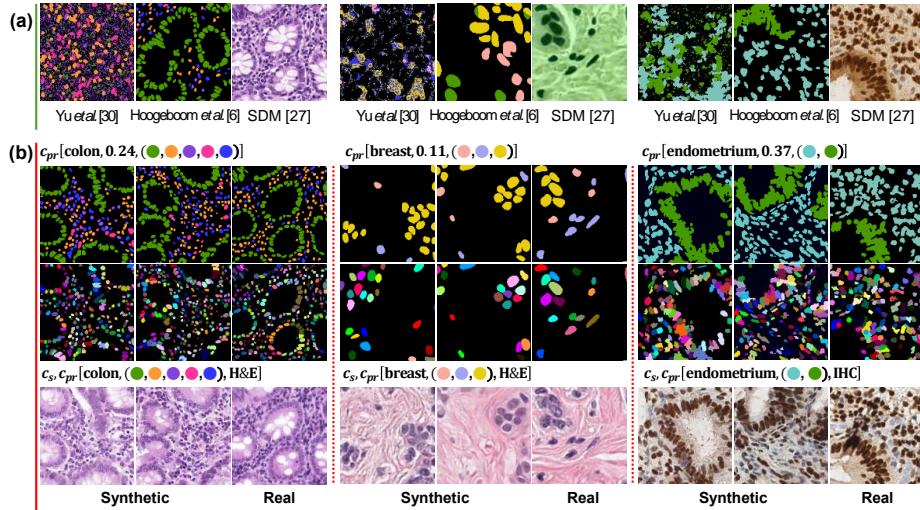
Fig. 2: Example of generated data using text conditions c_{pr} (b). (a) shows the synthetic data by related works.

Furthermore, we incorporate a pre-trained text encoder in the pathology domain, PLIP [29], encoding text prompts depicting tissue type, nuclei proportions, and class types while freezing the parameters of the text encoder. This allows fine control over the label generation process through the alignment of text and image (see Fig. 2), preventing the learning bias in the distribution of the training data that unconditional models have (see Fig. 3a). Conditioned by text embedding c_{pr}, the joint diffusion model is trained through the backward diffusion process $p_\theta(m_{t-1}|m_t, c_{pr})$ as depicted in Fig. 1a, enabling more sophisticated label generation using text-image alignments. Lastly, we create instance label using synthesized semantic label y_0 and distance masks x_0 by marker-controlled watershed algorithm [28].

2.3 Conditional Image Synthesis with Pretrained LDM

Latent diffusion model (LDM) [20] compresses an input image into lower dimensional vector z and performs the diffusion process within the latent space, lowering the cost required for training and inference of diffusion models. The generated sample z_0 is then reconstructed into a high resolution image in the pixel space. Since training an LDM from scratch consumes much resources, we fine-tune a text conditional LDM for pathology, PathLDM [29], to leverage a rich domain-specific knowledge without changing the pretrained weights. To synthesize the images reflecting semantic conditions, we introduce a scalable component to influence the generation process. Inspired by [32], the scalable component is a trainable copy of the diffusion model that is connected to the original model

via zero-initialized convolutional layers to prevent adversely affecting noise to the network's parameters. We only train this scalable component with semantic condition $c_s \in \mathbb{R}^{H \times W \times (K+3)}$, which is a combination of semantic labels, distance transform maps, and instance edge map. As illustrated in Fig. 1b, the semantic condition passes through learnable embedding layer, and computed with z_t. The training objective is then formulated as:

$$\mathbb{E}_{z,\epsilon \sim \mathcal{N}(0,1),t,c_s,c_{pr}} \left[\left\| \epsilon - \epsilon_\theta(z_t, c_s, c_{pr}, t) \right\|_2^2 \right], \tag{7}$$

where c_{pr} represents the text embedding.

3   Experiment

3.1   Datasets

We utilize three nuclei segmentation datasets: the first two, Lizard [8] and Pan-Nuke [6], were stained with hematoxylin and eosin (H&E), and the third, En-doNuke [15], was stained using immunohistochemistry (IHC) techniques. Lizard is the largest publicly available pathology nuclei dataset, consisting of colon tissue samples. It is derived from six different data sources and contains 495,179 nuclei, representing collections from multiple institutions. We pre-process the image samples into 13,064 256×256 sized patches as done by [23]. PanNuke is a pan-cancer nuclei dataset, incorporating 19 different tissue types. It features 205,343 nuclei across 7,901 patches, each 256×256 pixels in size at 40x higher magnification. Finally, EndoNuke is a endometrial tissue dataset, including 210,419 nuclei labels. These images have been resized and organized into a set of 1,780 patches, each 256×256 pixels in size. For training the generative model, we divided each dataset into training and test sets with these ratios: 80:20 for PanNuke, 85:15 for EndoNuke, and the NASDM [23] split for the Lizard dataset.

3.2   Implementation details

We use a single A6000 GPU with batch size 6 for the label synthesis model and a single RTX4090 GPU with batch size 16 for the image synthesis model. We set the timestep $T = 1000$ for training and $T = 100$ for DDIM-based sampling [24] for both models. For image and label synthesis, we utilize distinct text prompts. The label synthesis prompts are based on the ground truth labels and data descriptions. These prompts define the tissue type (e.g., colon), nucleus proportion that is ratio of pixels representing nuclei in the label map, and identify the types of nuclei present. Similarly, for image generation, the prompts include both tissue type and nuclei type, as used for label synthesis, and additionally specify the staining method (e.g., H&E or IHC). For the label synthesis model, we utilized the U-Net model from [16] and modified input/output channels as data pairs and followed text conditioning as done by Imagen [21].

Table 1: Quantitative results on generated labels using FSD metric and inference time per sample in second. [·] is type of text conditions: Tis, NP, and C indicates tissue type, nucleus proportion, and class types, respectively.

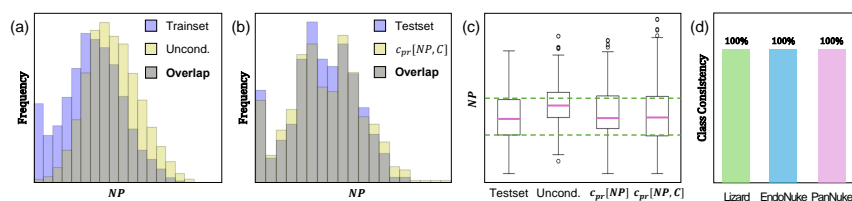| Method	Lizard	PanNuke	EndoNuke	Time (sec)
Yu *et al.* [30]	1018.19	1130.74	940.41	122.14
Hoogeboom *et al.* [12]	36.20	438.19	141.49	28.73
Ours (Uncond.)	12.00	41.28	190.81	-
$+ c_{pr}[Tis]$	-	9.11	-	-
$+ c_{pr}[Tis, NP]$	4.40	4.45	60.42	-
$+ c_{pr}[Tis, NP, C]$	**4.10**	**4.31**	**15.11**	**17.85**

Fig. 3: Graphical analysis. The unconditional model tends to replicate the high-frequency labels from the training data set (a), while the conditional model can generate labels with specific nucleus proportions (b). (c) shows that the nucleus proportion and class type conditions are effective for synthesizing the target label set. We also observe that it is 100% consistent for the given class types as (d).

3.3   Results

In Table 1, we demonstrate quantitative results on label synthesis comparing with Yu *et al.* [30], Hoogeboom *et al.* [12], and ablation studies of text conditions. First, we use Fréchet Segmentation Distance (FSD) [2] metric. The labels synthesized with Gaussian diffusion model by Yu *et al.* show a poor FSD score with random-looking noise (see Fig. 2a). The categorical diffusion model suggested by Hoogeboom *et al.* demonstrates a comparable FSD score on the Lizard dataset; however, generated label classes are biased. Similarly, our unconditional model also showed biased generation as depicted in Fig. 3a. As we incorporated more structure-related text conditions, our model well utilized the text-image alignment for desired label generation. Especially, tissue type text conditioning affects well in PanNuke, the multi-tissue type dataset.

Table 2 shows the image quality evaluation compared to semantic conditional generative methods using Fréchet Inception Distance (FID) [10] and Inception Score (IS) [22] metrics. Our fine-tuned semantic-text conditional LDM achieved the best in FID and second in IS next to SDM. However, as shown in Fig. 2, SDM generated unrealistic color images, which leads to higher IS. Moreover, as LDM requires less time for image generation, it can dramatically reduce the time for image generation, up to 33.6× faster, compared to SDM. As a result, the tailored

Table 2: Quantitative comparison of generated image quality to semantic condition-based generative models with FID and IS metrics, and required time to generate a sample in second.

Method	Lizard		PanNuke		EndoNuke		Time (sec)
	FID↓	IS↑	FID↓	IS ↑	FID↓	IS ↑	
pix2pixHD [26]	99.97	1.59	100.36	2.82	95.37	1.95	-
SPADE [19]	60.93	1.71	<u>71.03</u>	3.35	<u>68.02</u>	1.63	-
SDM [27]	<u>48.16</u>	**2.45**	95.44	**3.92**	91.42	**2.24**	180
Ours	**37.69**	<u>2.36</u>	**50.29**	<u>3.59</u>	**57.18**	<u>2.14</u>	5.35

Table 3: Downstream segmentation and classification performance. The best results are in **bold** and the second best are <u>underlined</u>.

Method	Lizard			PanNuke			Endonuke		
	Dice	AJI	F_d	Dice	AJI	F_d	Dice	AJI	F_d
Baseline	0.618	0.381	0.616	0.782	0.598	0.763	0.877	0.601	0.811
Conventional Aug.	0.675	0.423	0.646	0.816	0.641	0.791	0.891	0.624	0.823
CutOut [3]	0.700	0.463	0.674	0.799	0.631	0.781	<u>0.912</u>	0.649	0.843
CutMix [31]	0.690	0.444	0.661	0.807	0.643	0.787	0.898	0.586	0.799
SDM [27]	**0.718**	**0.487**	**0.699**	<u>0.821</u>	<u>0.654</u>	**0.800**	**0.914**	**0.661**	<u>0.847</u>
Ours	<u>0.714</u>	<u>0.479</u>	<u>0.692</u>	**0.822**	**0.655**	<u>0.798</u>	0.911	<u>0.656</u>	**0.850**

LDM can produce high-quality images that match the semantic and textual conditions, providing a computationally efficient alternative to traditional pixel-based diffusion models for pathology nuclei image synthesis.

We also conducted downstream nuclei segmentation and classification with synthetic data, using patches that were excluded from training the generative model. We use HoVer-Net [9] as our baseline network for downstream task evaluation. We utilize Dice coefficient and Aggregated Jaccard Index (AJI) [14] metrics to quantify segmentation performance, and detection quality F_d for classification. In Table 3, the conventional augmentation improves the baseline performance. We generated image-label data from a text prompt that conditions the label generation process and augmented the training set with the same number of patches as SDM's. For SDM data generation, we used ground truth labels. Ours improved baseline with conventional augmentation and comparable score within every metric to SDM. Considering the data generation cost of SDM as in Table 2, our approach is highly promising.

4   Conclusion

In this paper, we proposed a novel two-stage multi-class nuclei data augmentation framework, consisting of label synthesis followed by image synthesis. We showed that the text conditional label synthesis strategy enables manipulation of

the synthesized labels. To the best of our knowledge, this is the first attempt to leverage pretrained LDM for nuclei image synthesis aligning with the corresponding label, securing both high-quality sampled images and resource effectiveness. Lastly, we validate the efficacy of the proposed scheme on downstream tasks, improving the nuclei segmentation and classification performance.

For future work, we plan to explore synthesizing bigger-sized images, such as entire whole slide images, to expand the scalability of generative data augmentation.

Disclosure of Interests. There are no conflicts of interest to declare.

References

1. Alom, Z., Asari, V.K., Parwani, A., Taha, T.M.: Microscopic nuclei classification, segmentation, and detection with improved deep convolutional neural networks (dcnn). Diagnostic Pathology **17**(1), 1–17 (2022)
2. Bau, D., Zhu, J.Y., Wulff, J., Peebles, W., Strobelt, H., Zhou, B., Torralba, A.: Seeing what a gan cannot generate. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4502–4511 (2019)
3. DeVries, T., Taylor, G.W.: Improved regularization of convolutional neural networks with cutout. arXiv preprint arXiv:1708.04552 (2017)
4. Dhariwal, P., Nichol, A.: Diffusion models beat gans on image synthesis. Advances in neural information processing systems **34**, 8780–8794 (2021)
5. Doan, T.N.N., Kim, K., Song, B., Kwak, J.T.: Gradmix for nuclei segmentation and classification in imbalanced pathology image datasets. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 171–180. Springer (2022)
6. Gamper, J., Koohbanani, N.A., Benes, K., Graham, S., Jahanifar, M., Khurram, S.A., Azam, A., Hewitt, K., Rajpoot, N.: Pannuke dataset extension, insights and baselines. arXiv preprint arXiv:2003.10778 (2020)
7. Gong, X., Chen, S., Zhang, B., Doermann, D.: Style consistent image generation for nuclei instance segmentation. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 3994–4003 (2021)
8. Graham, S., Jahanifar, M., Azam, A., Nimir, M., Tsang, Y.W., Dodd, K., Hero, E., Sahota, H., Tank, A., Benes, K., et al.: Lizard: a large-scale dataset for colonic nuclear instance segmentation and classification. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 684–693 (2021)
9. Graham, S., Vu, Q.D., Raza, S.E.A., Azam, A., Tsang, Y.W., Kwak, J.T., Rajpoot, N.: Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. Medical image analysis **58**, 101563 (2019)
10. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium. Advances in neural information processing systems **30** (2017)

11. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)
12. Hoogeboom, E., Nielsen, D., Jaini, P., Forré, P., Welling, M.: Argmax flows and multinomial diffusion: Learning categorical distributions. Advances in Neural Information Processing Systems **34**, 12454–12465 (2021)
13. Ilyas, T., Mannan, Z.I., Khan, A., Azam, S., Kim, H., De Boer, F.: Tsfd-net: Tissue specific feature distillation network for nuclei segmentation and classification. Neural Networks **151**, 1–15 (2022)
14. Kumar, N., Verma, R., Sharma, S., Bhargava, S., Vahadane, A., Sethi, A.: A dataset and a technique for generalized nuclear segmentation for computational pathology. IEEE transactions on medical imaging **36**(7), 1550–1560 (2017)
15. Naumov, A., Ushakov, E., Ivanov, A., Midiber, K., Khovanskaya, T., Konyukova, A., Vishnyakova, P., Nora, S., Mikhaleva, L., Fatkhudinov, T., et al.: Endonuke: Nuclei detection dataset for estrogen and progesterone stained ihc endometrium scans. Data **7**(6), 75 (2022)
16. Nichol, A.Q., Dhariwal, P.: Improved denoising diffusion probabilistic models. In: International Conference on Machine Learning. pp. 8162–8171. PMLR (2021)
17. Oh, H.J., Jeong, W.K.: Diffmix: Diffusion model-based data synthesis for nuclei segmentation and classification in imbalanced pathology image datasets. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2023. pp. 337–345. Springer (2023)
18. Park, M., Yun, J., Choi, S., Choo, J.: Learning to generate semantic layouts for higher text-image correspondence in text-to-image synthesis. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 7591–7600 (2023)
19. Park, T., Liu, M.Y., Wang, T.C., Zhu, J.Y.: Semantic image synthesis with spatially-adaptive normalization. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 2337–2346 (2019)
20. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 10684–10695 (2022)
21. Saharia, C., Chan, W., Saxena, S., Li, L., Whang, J., Denton, E.L., Ghasemipour, K., Gontijo Lopes, R., Karagol Ayan, B., Salimans, T., et al.: Photorealistic text-to-image diffusion models with deep language understanding. Advances in neural information processing systems **35**, 36479–36494 (2022)
22. Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., Chen, X.: Improved techniques for training gans. Advances in neural information processing systems **29** (2016)
23. Shrivastava, A., Fletcher, P.T.: Nasdm: Nuclei-aware semantic histopathology image generation using diffusion models. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2023. pp. 786–796. Springer (2023)
24. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. arXiv preprint arXiv:2010.02502 (2020)
25. Wang, H., Xian, M., Vakanski, A., Shareef, B.: Sian: style-guided instance-adaptive normalization for multi-organ histopathology image synthesis. In: 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI). pp. 1–5. IEEE (2023)
26. Wang, T.C., Liu, M.Y., Zhu, J.Y., Tao, A., Kautz, J., Catanzaro, B.: 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 (2018)
27. Wang, W., Bao, J., Zhou, W., Chen, D., Chen, D., Yuan, L., Li, H.: Semantic image synthesis via diffusion models. arXiv preprint arXiv:2207.00050 (2022)

28. Yang, X., Li, H., Zhou, X.: Nuclei segmentation using marker-controlled watershed, tracking using mean-shift, and kalman filter in time-lapse microscopy. IEEE Transactions on Circuits and Systems I: Regular Papers **53**(11), 2405–2414 (2006)
29. Yellapragada, S., Graikos, A., Prasanna, P., Kurc, T., Saltz, J., Samaras, D.: Pathldm: Text conditioned latent diffusion model for histopathology. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 5182–5191 (2024)
30. Yu, X., Li, G., Lou, W., Liu, S., Wan, X., Chen, Y., Li, H.: Diffusion-based data augmentation for nuclei image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 592–602. Springer (2023)
31. Yun, S., Han, D., Oh, S.J., Chun, S., Choe, J., Yoo, Y.: Cutmix: Regularization strategy to train strong classifiers with localizable features. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) (October 2019)
32. Zhang, L., Rao, A., Agrawala, M.: Adding conditional control to text-to-image diffusion models. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3836–3847 (2023)