Multimodal-Conditioned Flow Matching for Pathology Nuclei Data Augmentation

Yanan Zhang[1], Yinuo Wang[1], Mengyang Liu[1], and Xiangzhi Bai[1,2,3](✉)

[1] Image Processing Center, School of Astronautics, Beihang University, Beijing, China
[2] State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing, China
[3] Key Laboratory of Spacecraft Design Optimization and Dynamic Simulation Technology, Ministry of Education, Beihang University, Beijing, China
jackybxz@buaa.edu.cn

Abstract. Nuclei segmentation and classification are fundamental to computational pathology, yet the performance of supervised deep learning approaches is often hampered by the scarcity of large-scale, accurately annotated datasets. Although generative data augmentation offers a promising solution, existing methods remain constrained by inefficient training and limited efficacy on small-scale annotated datasets. To address these issues, we propose a novel multimodal conditioned framework based on optimal transport flow matching (OTFM) to synthesize high-fidelity nuclei mask and pathology image pairs. Our framework employs synchronous multimodal conditioning to efficiently utilize small-scale datasets for generating aligned nuclei mask-image pairs, and adopts few-step sampling to accelerate inference. To ensure biologically plausible mask generation, we integrate cell center points and text descriptions as conditions to preserve spatial layout consistency. Furthermore, self-supervised learning (SSL) feature representations extracted from pathology foundation models are incorporated as additional conditions, providing morphological guidance for synthesizing high-fidelity, accurately aligned pathology images. Extensive experiments on pathology nuclei datasets show that our framework consistently improves performance in nuclei segmentation and classification tasks. Code is available at https://github.com/zhangyn1415/MMCFM-PNDA.

Keywords: Computational pathology · Flow matching models · Foundation models · Nuclei segmentation and classification · Data augmentation.

1 Introduction

The rapid advancement of computational pathology (CPath) has facilitated the widespread adoption of deep learning (DL) models in pathological diagnosis. Nuclei segmentation and classification serve as fundamental steps for tissue analysis, providing critical information on cellular architecture and enabling quantitative

studies of disease pathology [22, 20, 6]. However, fully supervised DL methods remain constrained by the scarcity of large-scale, accurately annotated datasets due to the labor-intensive nature of manual nucleus annotation [14, 11]. Generative data augmentation, which employs generative models to synthesize realistic pathology images, has emerged as an effective strategy to enhance DL model performance across medical imaging tasks [12]. Recently, flow matching (FM) has emerged as an efficient generative modeling paradigm, offering improved convergence properties and sampling efficiency compared to traditional score-based diffusion approaches by establishing straight probability paths between distributions [13].

In nuclei segmentation and classification, existing generative augmentation methods [24, 27, 21] typically rely on nuclei masks as conditions, resulting in limited controllability and realism, as pathology images exhibit organ and tissue-specific textural characteristics not fully captured by nuclei masks alone [15]. Recent works incorporate text descriptions as additional conditions with diffusion frameworks, yet they suffer from inefficient multi-stage training and require numerous sampling steps [18, 1, 23]. Moreover, text descriptions alone provide insufficient guidance for capturing detailed morphological and textural features of pathology images.

Computational pathology foundation models (CPathFMs), pretrained on large-scale unlabeled histopathology data via self-supervised learning (SSL), demonstrate strong generalization capabilities [3, 17]. SSL features extracted from CPathFMs preserve rich visual characteristics, serving as effective conditional guidance for high-fidelity pathology image generation [9].

In this paper, we propose a novel multimodal conditioned flow matching framework for synthesizing high-fidelity, accurately aligned nuclei mask and pathology image pairs. Specifically, we adopt synchronous multimodal conditioning based on optimal transport flow matching (OTFM), enabling end-to-end training and few-step sampling to improve both training and inference efficiency. For nuclei mask synthesis, we integrate cell center points and textual descriptions as conditions to preserve spatially consistent layouts aligned with pathological knowledge. For pathology image synthesis, we further incorporate SSL feature representations from CPathFMs to provide fine-grained morphological and textural guidance. Extensive experiments on multiple nuclei segmentation and classification datasets demonstrate that our synthetic data substantially improves downstream task performance. The main contributions are:

(1) A novel multimodal conditioned flow matching framework for synthesizing paired nuclei masks and pathology images, effectively enhancing DL model performance in nuclei segmentation and classification tasks.

(2) Integration of multimodal conditions combining SSL feature representations from CPathFMs and text embeddings from pathology vision-language models (VLMs) to guide high-fidelity pathology image synthesis.

(3) Synchronous multimodal conditioning based on OTFM that achieves improved training and inference efficiency through end-to-end training and few-step sampling.

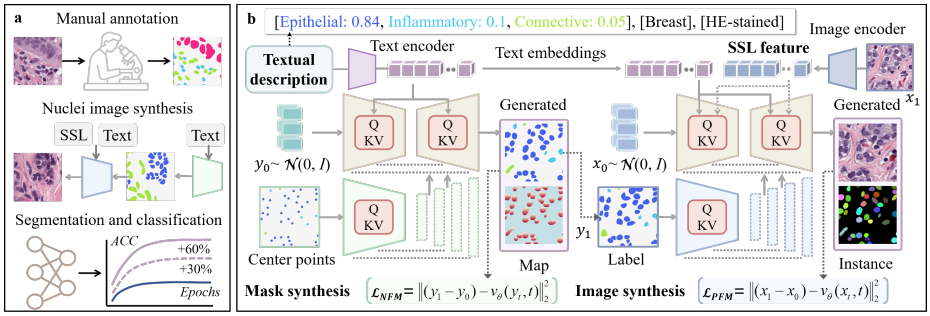

Fig. 1. Architecture of our multimodal conditioned flow matching framework for synthesizing pathology image-mask pairs.

2 Method

In this section, we present our framework for synthesizing paired nuclei masks and pathology images, as illustrated in Fig. 1. The pathology nuclei datasets consists of pathology image $x \in \mathbb{R}^{H \times W \times 3}$ and nuclei mask $y \in \mathbb{R}^{H \times W \times K}$, where K represents the number of nuclei types. The following sections detail our approach: Section 2.1 introduces flow matching, Section 2.2 describes pathology image synthesis, and Section 2.3 presents nuclei mask synthesis.

2.1 Flow matching models

Flow matching models generate data by deterministically transforming samples from a simple Gaussian distribution to a target data distribution. In contrast to diffusion models that rely on stochastic differential equations (SDEs), FM models establishes a straight-line trajectory between the noise sample x_0 and data sample x_1, parameterized as:

$$x_t = (1 - t)x_0 + tx_1, \quad t \in [0, 1]. \tag{1}$$

We adopt optimal transport flow matching, which pairs noise and data samples via optimal transport coupling to establish straight probability paths. The OTFM loss is defined as:

$$\mathcal{L}_{\text{OTFM}} = \mathbb{E}_{(x_0, x_1)} \left[\|(x_1 - x_0) - v_\theta(x_t, t)\|_2^2 \right], \tag{2}$$

During inference, data samples are synthesized by integrating the learned vector field:

$$x_1 = x_0 + \int v_\theta(x_\tau, \tau) d\tau, \tag{3}$$

which corresponds to solving the ordinary differential equation $dx_t/dt = v_\theta(x_t, t)$. As illustrated in Fig. 2(a), this formulation enables image synthesis with fewer sampling steps compared to diffusion-based approaches.

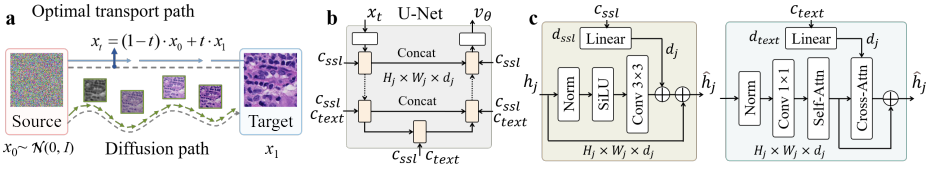

Fig. 2. (a) Illustration of sampling process. (b) Illustration of the U-Net architecture. (c) SSL feature and text embedding fusion modules.

2.2 Multimodal conditional pathology image synthesis

We synthesize pathology images by conditioning the flow matching model on nuclei type maps c_m, text embeddings c_{text}, and SSL features $c_{\text{ssl}} = \mathcal{E}_{\text{ssl}}(x) \in \mathbb{R}^{d_{\text{ssl}}}$ extracted from a pathology foundation model $\mathcal{E}_{\text{ssl}}$. These conditions provide complementary structural, semantic, and morphological guidance for high-fidelity pathology image generation.

As shown in Fig. 1(b), the nuclei type maps $c_m = y_{\text{cls}} \in \mathbb{R}^{H\times W\times K}$ derived from nuclei mask label y, are encoded by an auxiliary U-Net branch and injected into the decoder via zero convolutions, ensuring that synthesized pathology images retain accurate nuclei structures and type information. Textual descriptions provide complementary semantic information, specifying organ tissue types and nuclei class proportions. We integrate text embeddings $c_{\text{text}} \in \mathbb{R}^{d_{\text{text}}}$ extracted from a pathology VLM via cross-attention in the deeper U-Net layers, where the reduced spatial resolution enables efficient computation and the abstract feature representations naturally align with the semantic content of the text (Fig. 2).

SSL features encode fine-grained morphological and textural priors that guide generative models toward realistic pathology image synthesis [9]. We introduce an SSL feature fusion module (SFFM) to incorporate these features across all U-Net layers (Fig. 2(c)). For feature maps h_j at layer j, SFFM computes:

$$\tilde{h}_j = W_d \cdot \text{SiLU}(\text{Norm}(h_j)), \quad \hat{h}_j = \tilde{h}_j + W_l c_{\text{ssl}}, \tag{4}$$

where W_d is a 3×3 convolution, $\text{Norm}(\cdot)$ is group normalization, and W_l is a linear projection. By integrating SFFM into every U-Net layer, SSL features provide multi-scale guidance that captures both texture and morphological structure, enabling high-fidelity pathology image synthesis.

Our framework enables end-to-end synchronous training of multimodal conditions, leveraging the training efficiency of OTFM to achieve faster convergence than multi-stage diffusion-based approaches. The training objective is defined as:

$$\mathcal{L}_{\text{PFM}} = \mathbb{E}_{x_0, x_1, c_{\text{ssl}}, c_{\text{text}}, c_m} \left[\| (x_1 - x_0) - v_\theta(x_t, t, c_{\text{ssl}}, c_{\text{text}}, c_m) \|_2^2 \right]. \tag{5}$$

2.3 Multi-class nuclei mask label synthesis

We synthesize multi-class nuclei masks by conditioning the FM model on text embeddings c_{text} and cell center points c_p derived from nuclei mask label y. As

shown in Fig. 1(b), these conditions provide complementary semantic and geometric guidance for generating nuclei masks with accurate cellular arrangements.

From a multi-class nuclei mask y, we extract the nuclei type map $y_{\text{cls}} = \text{clamp}(y, 0, 1)$ and instance map $y_{\text{inst}} = \max_k y_{[:,:,k]}$. The cellular spatial layout is represented by cell center points $c_p \in \mathbb{R}^{H\times W\times K}$ and encoded by an auxiliary U-Net branch with zero-convolution injection to preserve geometric fidelity.

To further guide instance-level nuclei synthesis, we derive additional spatial representations: a binary nuclei mask $\bar{y} \in \mathbb{R}^{H\times W\times 1}$ and horizontal-vertical distance maps $y_{\text{hv}} \in \mathbb{R}^{H\times W\times 2}$. The synthesized nuclei mask is formed as: $y' = [y_{\text{cls}}, \bar{y}, y_{\text{hv}}]$, and post-processed via watershed segmentation to obtain instance-aware results: $y'_{\text{inst}} = \text{watershed}(c_p, \bar{y}, y_{\text{hv}})$, yielding the final instance-aware multi-class nuclei mask label $(y'_{\text{cls}}, y'_{\text{inst}})$.

As illustrated in Fig. 2, we preserve the U-Net architecture and text embeddings c_{text} are used as conditions. The training objective combines velocity field prediction with multimodal conditioning:

$$\mathcal{L}_{\text{NFM}} = \mathbb{E}_{y_0, y_1, c_{\text{text}}, c_p}\left[\|(y_1 - y_0) - v_\theta(y_t, t, c_{\text{text}}, c_p)\|_2^2\right], \tag{6}$$

where $y_0 \sim \mathcal{N}(0, I)$ is a noise sample, y_1 denotes the target mask y', and $y_t = (1-t)y_0 + ty_1$ follows the FM formulation.

3 Experiment and Results

3.1 Pathology nuclei datasets

PanNuke [5] contains 189,744 annotated nuclei across 19 organ types, categorized into five classes; we adopt the standard three-fold split, using the first two folds for training and the last for test. **EndoNuke** [16] comprises 1,780 endometrial immunohistochemistry (IHC) image patches with 215,758 nuclei across three categories, using a 60%/40% split. **Lizard** [7] provides nearly one million annotated nuclei from colon whole-slide images (WSIs) across five sub-datasets, with patches from three sub-datasets used for training and the remaining two for test.

3.2 Implementation details

Our framework was trained on 256×256 image patches for 150 epochs with a batch size of 4 and learning rate of 10^{-4} using an NVIDIA RTX 4090 GPU. Our U-Net consists of five resolution levels, with text embeddings integrated into the three lowest-resolution levels. In downstream tasks, we trained CellViT [10] with a batch size of 16 for 100 epochs, freezing the first 15 epochs and using a learning rate of 0.003 while keeping other parameters default. SSL features were extracted using the pretrained image encoder UNI [2] for PanNuke and EndoNuke, and Virchow2 [29] for Lizard. Text descriptions were encoded via the text encoder of MUSK [25].

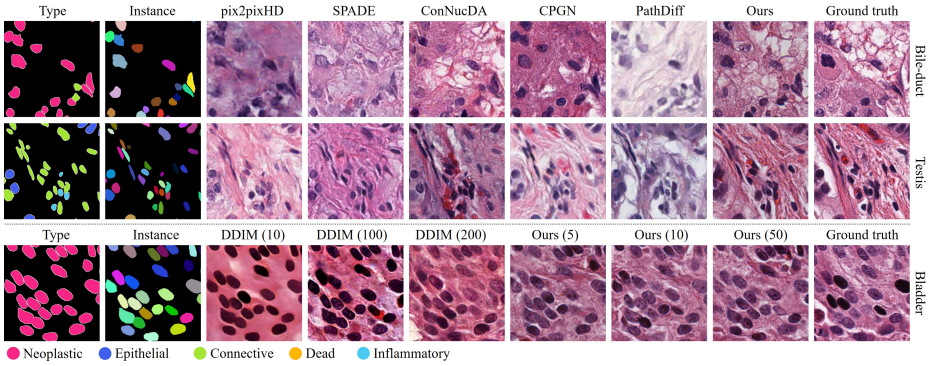

Fig. 3. Visual comparison of image generation performance with PanNuke.

Table 1. Quantitative evaluation of image generation quality with PanNuke.

Method	pix2pixHD	SPADE	Nudiff	ConNucDA	CPGN	PathDiff	Ours
CLIP-FID	8.12	8.72	20.15	7.94	5.46	10.36	**4.51**
KID	92.73	93.06	116.83	65.66	57.62	67.17	**50.91**
LPIPS	49.76	49.05	58.53	48.45	47.75	49.45	**47.11**
IS(↑)	2.82	2.95	2.97	3.06	3.05	2.92	**3.35**
mDice(↑)	46.65	46.28	45.58	46.52	47.15	45.85	**47.58**
mIoU(↑)	43.85	43.65	43.28	44.12	45.05	43.52	**45.57**

3.3 Image generation quality

We compare our method against GAN-based [24, 19, 28] and diffusion-based [27, 18, 1] generative approaches. While existing methods incorporate text descriptions as additional conditions, they are often limited by insufficient conditioning information or inefficient multi-stage training, resulting in uncertain morphological and textural features in synthesized pathology images. In contrast, our framework integrates SSL features with end-to-end synchronous training of multimodal conditions, enabling more accurate synthesis of tissue-specific textures and morphologies. As shown in Fig. 3, our generated images better preserve nuclei morphology and tissue characteristics.

Quantitative comparison For fair evaluation, we use pathology images synthesized by different methods as additional training data for the nuclei segmentation model HoVerNet [8]. Table 1 presents the comparison results measured by segmentation metrics (mDice, mIoU) and image quality metrics (IS, CLIP-FID, KID, LPIPS). Our method consistently outperforms baselines across all metrics, indicating more accurate synthesis of multi-class nuclei morphology and spatial distributions.

Inference efficiency As shown in Fig. 3, our framework generates pathology images with few-step sampling, whereas 10-step and 100-step DDIM sampling [18] generate blurry images lacking cellular detail. Table 2 compares synthesis

quality across different sampling steps, demonstrating the efficiency advantage of our approach for pathology image generation.

Table 2. Ablation study with PanNuke.

c_{ssl}	c_{text}	Steps	CLIP-FID	KID	LPIPS	IS(↑)	mDice(↑)	mIoU(↑)
×	×	10	20.02	104.84	55.28	2.58	42.42	31.95
×	✓	10	5.68	58.65	48.21	3.06	44.38	42.16
✓	×	10	5.13	53.45	47.74	3.16	44.82	42.75
✓	✓	5	4.88	**50.31**	<u>47.19</u>	3.32	**45.32**	43.45
✓	✓	10	**4.51**	50.91	**47.11**	<u>3.35</u>	<u>45.29</u>	**43.51**
✓	✓	50	<u>4.67</u>	<u>50.58</u>	47.25	**3.39**	45.01	<u>43.39</u>

Ablation study We conduct an ablation study on PanNuke to assess the contribution of each framework component (Table 2). Our framework fully integrates multimodal conditions to improve semantic alignment with tissue descriptions and enhance morphological details. Combined with nuclei masks, these conditions yield precise structural correspondence and high-fidelity pathology images.

Table 3. Effectiveness of the proposed data augmentation method with CellViT.

Datset	PanNuke			EndoNuke			Lizard		
Split	Baseline	+30%	+60%	Baseline	+30%	+60%	Baseline	+30%	+60%
Dice	81.57	82.56	**82.72**	93.08	93.34	**93.46**	74.21	74.85	**75.13**
IoU	70.43	71.98	**72.08**	87.15	87.65	**87.92**	60.50	61.61	**61.67**
mPQ	44.49	45.57	**45.64**	58.04	**58.64**	58.53	26.59	**27.54**	27.48
F1	80.10	80.25	**80.38**	84.44	84.81	**84.91**	68.38	68.89	**69.04**

Downstream tasks Fig. 4 shows examples of synthesized nuclei mask and pathology image pairs. We evaluate our data augmentation approach using CellViT as the baseline model. The original training set is augmented with synthetic mask-image pairs at 30% and 60% ratios. Performance is measured using Dice and IoU for segmentation, and mean Panoptic Quality (mPQ) and F1-score for classification and detection. Table 3 shows consistent improvements across diverse pathology nuclei datasets, validating the effectiveness of our data augmentation method.

Visualization Fig. 5(a) shows that synthetic masks from our method closely align with real mask distributions. Quantitative morphological comparisons (area, perimeter, eccentricity, solidity) further confirm this alignment, with all relative

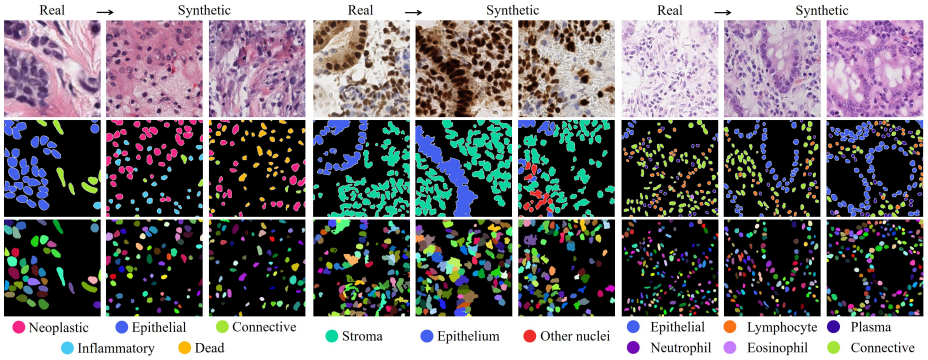

Fig. 4. Samples of synthesized pathology images and nuclei mask labels.

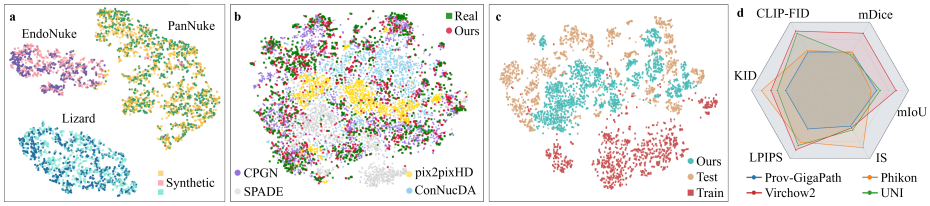

Fig. 5. Visualization of feature distributions using t-SNE. (a) Multi-class nuclei masks across datasets. (b) Generated pathology images from PanNuke using different methods. (c) Pathology images from Lizard. (d) Comparison of different CPathFMs.

differences below 3% across the three datasets. For PanNuke, real features exhibit dispersed distributions reflecting organ and tissue diversity (Fig. 5(b)). Our synthetic images better approximate this dispersion, enhancing visual realism. This improvement is attributed to the integration of SSL features, which provide accurate morphological and textural priors. For Lizard, where training and test samples originate from different sub-datasets, Fig. 5(c) shows that our synthesized images approach the distribution of unseen test data, validating the generalization and domain adaptation capabilities of CPathFMs. We further validate the performance of diverse CPathFMs in Lizard and results in Fig. 5(d) indicate Virchow2 slightly outperforms other models [4, 26].

4 Conclusion

We propose a multimodal conditioned flow matching framework for synthesizing paired nuclei masks and pathology images. To generate high-fidelity pathology images with accurate morphological and textural details, we integrate SSL features from pathology foundation models and text embeddings as additional conditions. For nuclei mask synthesis, we incorporate cell center points to preserve spatially consistent layouts aligned with pathological knowledge. Our framework

adopts synchronous multimodal conditioning based on OTFM to improve training efficiency, and leverages few-step sampling for accelerated inference. Experimental results on datasets encompassing diverse cell types and staining protocols confirm that our method consistently improves the performance of DL models in pathology nuclei segmentation and classification tasks.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (Grant No.62271016), and the Beijing Natural Science Foundation (Grant No. L242130).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Bhosale, M., Wasi, A., Zhai, Y., Tian, Y., Border, S., et al.: Pathdiff: Histopathology image synthesis with unpaired text and mask conditions. ArXiv preprint arXiv:2506.23440 (2025)
2. Chen, R., Ding, T., Lu, M., Williamson, D., Jaume, G., et al.: Towards a general-purpose foundation model for computational pathology. Nature Medicine **30**(3), 850–862 (2024)
3. Du, R., Carbonell, E., Huang, J., Liu, S., Wang, X., et al.: Ethics of foundation models in computational pathology: overview of contemporary issues and future implications. IEEE Transactions on Medical Imaging **44**(10), 4098–4115 (2025)
4. Filiot, A., Ghermi, R., Olivier, A., Jacob, P., Fidon, L., Camara, A., et al.: Scaling self-supervised learning for histopathology with masked image modeling. MedRxiv (2023)
5. Gamper, J., Koohbanani, N.A., Benes, K., Graham, S., Jahanifar, M., et al.: Pannuke dataset extension, insights and baselines. ArXiv preprint arXiv:2003.10778 (2020)
6. Gao, R., Yuan, X., Ma, Y., Wei, T., Johnston, L., et al.: Harnessing tme depicted by histological images to improve cancer prognosis through a deep learning system. Cell Reports Medicine **5**(5), 101536 (2024)
7. Graham, S., Jahanifar, M., Azam, A., Nimir, M., Tsang, Y., 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 (ICCV). pp. 684–693 (2021)
8. Graham, S., Vu, Q., Raza, S.A., Azam, A., Tsang, Y., et al.: Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. Medical Image Analysis **58**, 101563 (2019)
9. Graikos, A., Yellapragada, S., Le, M.Q., Kapse, S., Prasanna, P., et al.: Learned representation-guided diffusion models for large-image generation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 8532–8542 (2024)
10. Hörst, F., Rempe, M., Heine, L., Seibold, C., Keyl, J., , et al.: Cellvit: Vision transformers for precise cell segmentation and classification. Medical Image Analysis **94**, 103143 (2024)

11. Hou, L., Gupta, R., Van Arnam, J.S., Zhang, Y., Sivalenka, K., et al.: Dataset of segmented nuclei in hematoxylin and eosin stained histopathology images of ten cancer types. Scientific Data **7**(1), 185 (2020)
12. Kazerouni, A., Aghdam, E.K., Heidari, M., Azad, R., Fayyaz, M., et al.: Diffusion models in medical imaging: A comprehensive survey. Medical Image Analysis **88**, 102846 (2023)
13. Kornilov, N., Mokrov, P., Gasnikov, A., Korotin, A.: Optimal flow matching: Learning straight trajectories in just one step. Advances in Neural Information Processing Systems (NeruIPS) **37**, 104180–104204 (2024)
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. Mahmood, F., Borders, D., Chen, R.J., McKay, G.N., Salimian, K.J., et al.: Deep adversarial training for multi-organ nuclei segmentation in histopathology images. IEEE Transactions on Medical Imaging **39**(11), 3257–3267 (2019)
16. Naumov, A., Ushakov, E., Ivanov, A., Midiber, K., Khovanskaya, T., et al.: Endonuke: Nuclei detection dataset for estrogen and progesterone stained ihc endometrium scans. Data **7**(6), 75 (2022)
17. Neidlinger, P., El Nahhas, O.S., Muti, H.S., Lenz, T., Hoffmeister, M., et al.: Benchmarking foundation models as feature extractors for weakly supervised computational pathology. Nature Biomedical Engineering **9**(10), 1–11 (2025)
18. Oh, H., Jeong, W.: Controllable and efficient multi-class pathology nuclei data augmentation using text-conditioned diffusion models. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). pp. 36–46. Springer (2024)
19. Park, T., Liu, M., Wang, T., Zhu, J.: Semantic image synthesis with spatially-adaptive normalization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2337–2346 (2019)
20. Rong, R., Sheng, H., Jin, K.W., Wu, F., Luo, D., et al.: A deep learning approach for histology-based nucleus segmentation and tumor microenvironment characterization. Modern Pathology **36**(8), 100196 (2023)
21. Shrivastava, A., Fletcher, P.T.: Nasdm: Nuclei-aware semantic histopathology image generation using diffusion models. In: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). pp. 786–796. Springer (2023)
22. Stanton, S., Disis, M.: Clinical significance of tumor-infiltrating lymphocytes in breast cancer. Journal for Immunotherapy of Cancer **4**(1), 59 (2016)
23. Susladkar, O., Deshmukh, G., Tur, Y., Durak, G., Bagci, U.: Victr: Vital consistency transfer for pathology aware image synthesis. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 22772–22782 (2025)
24. Wang, T., Liu, M., Zhu, J., 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 (CVPR). pp. 8798–8807 (2018)
25. Xiang, J., Wang, X., Zhang, X., Xi, Y., Eweje, F., Chen, Y., et al.: A vision–language foundation model for precision oncology. Nature **638**(8051), 769–778 (2025)
26. Xu, H., Usuyama, N., Bagga, J., Zhang, S., Rao, R., Naumann, T., Wong, C., Gero, Z., et al.: A whole-slide foundation model for digital pathology from real-world data. Nature **630**(8015), 181–188 (2024)

27. 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 (MICCAI). pp. 592–602. Springer (2023)
28. Zhang, Y., Liu, Q., Chen, Q., Bai, X.: Clip-guided generative network for pathology nuclei image augmentation. Medical Image Analysis p. 103908 (2025)
29. Zimmermann, E., Vorontsov, E., Viret, J., Casson, A., Zelechowski, M., et al.: Virchow2: Scaling self-supervised mixed magnification models in pathology. arXiv preprint arXiv:2408.00738 (2024)