PathoPainter: Augmenting Histopathology Segmentation via Tumor-aware Inpainting

Hong Liu[*,1,†], Haosen Yang[*,2], Evi M.C. Huijben[1], Mark Schuiveling[3], Ruisheng Su[1], Josien P.W. Pluim[1], and Mitko Veta[1]

[1] Eindhoven University of Technology, Eindhoven, The Netherlands
[2] University of Surrey, London, UK
[3] University Medical Centre Utrecht, Utrecht, The Netherlands
* Joint main authors, † corresponding author: `hong.liu.00408@gmail.com`

Abstract. Tumor segmentation plays a critical role in histopathology, but it requires costly, fine-grained image-mask pairs annotated by pathologists. Thus, synthesizing histopathology data to expand the dataset is highly desirable. Previous works suffer from inaccuracies and limited diversity in image-mask pairs, both of which affect training segmentation, particularly in small-scale datasets and the inherently complex nature of histopathology images. To address this challenge, we propose **PathoPainter**, which reformulates image-mask pair generation as a tumor inpainting task. Specifically, our approach preserves the background while inpainting the tumor region, ensuring precise alignment between the generated image and its corresponding mask. To enhance dataset diversity while maintaining biological plausibility, we incorporate a sampling mechanism that conditions tumor inpainting on regional embeddings from a different image. Additionally, we introduce a filtering strategy to exclude uncertain synthetic regions, further improving the quality of the generated data. Our comprehensive evaluation spans multiple datasets featuring diverse tumor types and various training data scales. As a result, segmentation improved significantly with our synthetic data, surpassing existing segmentation data synthesis approaches, *e.g.*, 75.69% → 77.69% on CAMELYON16. The code is available at https://github.com/HongLiuuuuu/PathoPainter.

Keywords: Diffusion models · Histopathology segmentation · Data augmentation.

1 Introduction

Tumor segmentation is crucial in histopathology [11,24,21,15,7], but training effective segmentation models, especially in supervised scenarios, requires pixel-wise annotations on whole slide images (WSIs), which demands expert knowledge and significant effort due to their structural complexity and gigapixel scale. This often results in sparse and limited labeled datasets [1], as annotation of a single slide at high magnification can take an experienced pathologist hours. This
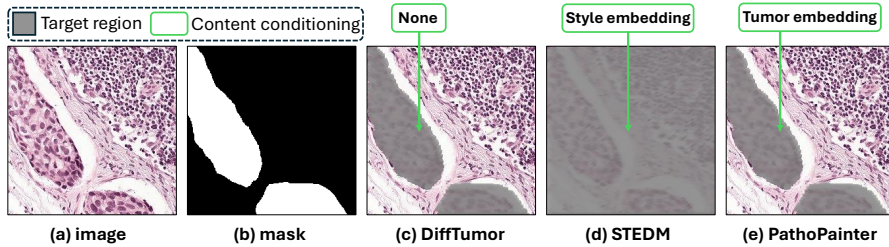
Fig. 1: Illustration of various synthetic methods. The columns represent: (a) the original image x, (b) the ground truth mask m, (c) DiffTumor [3], which inpaints the tumor region using the mask as a condition without any content conditioning, (d) STEDM [14], which generates the entire image with style conditioning, and (e) PathoPainter (ours), which inpaints the tumor region using the mask and tumor embeddings from a different image as conditions.

challenge, along with privacy regulations and ethical concerns [17,20], has led to the exploration of medical image-mask pair synthesis.

Recent advancements in generative models, particularly denoising diffusion probabilistic models (DDPMs) [9], have opened new avenues for conditional image synthesis [25,26]. DiffTumor [3] (Figure 1(c)) augments computed tomography data via inpainting. However it tends to overfit in histopathology data due to the absence of semantic conditioning, failing to capture the intricate variations in appearance and structure. Another approach [14] (Figure 1(d)) generates images conditioned on various styles to enhance dataset diversity. However, this method often cause misalignment between the generated image and the ground truth mask, resulting in inaccuracies. These issues negatively affect the training of segmentation models, which rely on high-quality image-mask pairs for optimal performance. Specifically in WSIs, the presence of subtle tissue patterns, intricate cellular details, and varying staining intensities make it crucial to maintain precise alignment between the image and its corresponding mask to ensure accurate segmentation [21,11,24,15].

In this paper, we propose **PathoPainter** (Figure 1(e)), which reformulates image-mask pair generation by inpainting images conditioned on regional tumor embeddings and masks. Specifically, we apply a mask to retain non-tumor features and use regional tumor embeddings as content conditioning to generate the tumor region during training. During inference, we employ a sampling mechanism that selects tumor embeddings from a different image to generate the tumor region and maintain biological plausibility. Synthetic image-mask pairs are used to train a segmentation model, where local inaccuracies, referred to as uncertain regions, can negatively impact learning. We hypothesize that uncertain regions in synthetic images can be identified by their higher loss values when evaluated using a segmentation model pretrained on real data. To mitigate their impact, we propose a self-adaptive approach to filter out these regions from the synthetic images. We summarize the contributions of this work as follows:
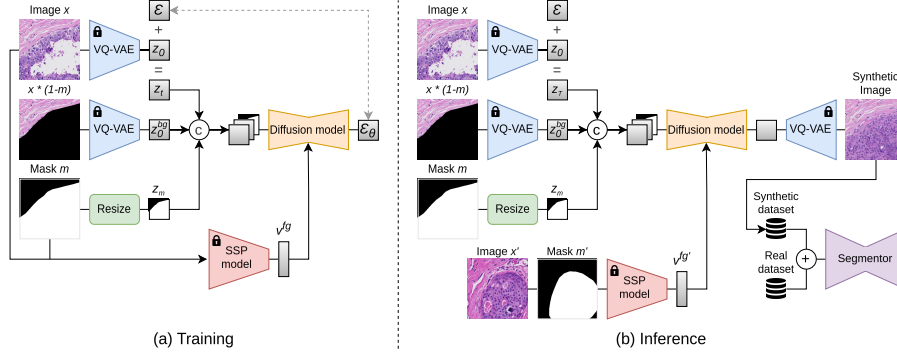
Fig. 2: Overview of the **PathoPainter** framework. During (a) training, the input image x and the masked image $x \odot (1-m)$ are encoded using a pretrained VQ-VAE [16] to obtain their latent representations z_0 and z_0^{bg}. These representations, concatenated with the resized mask z_m, are fed into the diffusion model and conditioned on the embedding v^{fg} extracted by SSP model, which captures features exclusively from the foreground. To augment the training set for downstream segmentation tasks, we generate reliable image-mask pairs using the method in (b). We randomly select a foreground embedding $v^{fg'}$ from the same tumor type as the target image. The U-Net denoiser then generates a new image-mask pair that contains tumors with characteristics distinct from those in the original image.

- We propose **PathoPainter**, which reformulates image-mask pair generation by inpainting images conditioned on masks and regional tumor embeddings.
- We introduce a conditioning mechanism that leverages regional tumor embeddings for content conditioning, combined with tumor embedding sampling and an uncertain region filtering strategy, to generate high-quality, accurate, and diverse image-mask pairs.
- We thoroughly evaluate the effectiveness of **PathoPainter** in enhancing dataset diversity and boosting segmentation performance.

2   Method

We propose a conditional diffusion model for histopathology image-mask pair synthesis (Figure 2). First, we review DDPM (Section 2.1), then introduce our Tumor-aware Conditional Stable Diffusion model: **PathoPainter** (Section 2.2), and finally, describe the filtering mechanism to address uncertain regions (Section 2.3).

2.1   Preliminaries: Conditional Diffusion model

The DDPM [9] is a generative framework designed to learn a target data distribution. To overcome the computational challenges of training DDPMs on high-

dimensional data, such as images, LDM [18] was introduced, enabling DDPM training in a latent space rather than the pixel space. LDMs consist of three key components: 1) a Variational Autoencoder (VAE) [10] that encodes an input image x into compact latent space z, 2) a U-Net denoiser that learns to reverse the diffusion process by iteratively refining the noisy latent representations over T time steps, and 3) a conditioning mechanism to guide and control the generation process. The U-Net denoiser is trained to predict the Gaussian noise ϵ added at each time step. Specifically, given the noisy latent z_t and a conditioning input c, it learns to minimize the difference between the true noise ϵ and the predicted noise $\epsilon_\theta(z_t, c, t)$. The LDM training objective is defined as

$$\mathcal{L}_{ldm} = \mathbb{E}_{t \sim \mathcal{U}(0,T), \epsilon \sim \mathcal{N}(0,1)} \left[\| \epsilon - \epsilon_\theta \left(z_t, c, t \right) \|_2^2 \right]. \tag{1}$$

2.2   PathoPainter

Training We first encode the input image x with a pretrained Vector quantized (VQ)-VAE [16] to obtain its latent representation z_0. Next, we encode the masked image $x \odot (1 - m)$ to derive a latent representation z_0^{bg} that reflects only background, or non-tumor, features. For compatibility, we resize the tumor mask m to match the dimensions of the latent representations and denote this resized mask as z_m.

The input to the diffusion model is the concatenation of z_t, z_0^{bg}, and z_m, where z_t is the initial latent representation z_0 with Gaussian noise added t times. Specifically, by inputting the original image x and the tumor mask m to the self-supervised pretrained (SSP) model [4], we extract a one-dimensional embedding v^{fg} with mask pooling technique [8]. This embedding represents regional tumor embedding, capturing information solely from the foreground (tumor) regions, as illustrated in the lower part of Figure 2. This process leads to a training objective that minimizes the difference between the added Gaussian noise ϵ and the predicted noise ϵ_θ for a time step t. The objective is defined as:

$$\mathcal{L}_{PathoPainter} = \mathbb{E}_{t \sim \mathcal{U}(0,T), \epsilon \sim \mathcal{N}(0,1)} \left[\left\| \epsilon - \epsilon_\theta \left(z_t, z_0^{bg}, v^{fg}, m, t \right) \right\|_2^2 \right]. \tag{2}$$

Inference To augment the training set, we generate reliable image-mask pairs (Figure 2(b)). Instead of using embeddings from the same image, our regional embedding sampling mechanism selects tumor embeddings $v^{fg'}$ from a different image within the same tumor group. To ensure biological plausibility, we cluster all tumor embeddings from the training set into k groups using K-means [13]. Given an image x for inpainting and another image x' providing $v^{fg'}$ as content conditioning, this approach enhances the diversity of the dataset.

Discussion Compared to directly generating an full image (Figure 1(d)), our approach results in more realistic and biologically plausible tumor regions that seamlessly blend into their surroundings. Additionally, by focusing on tumor regions within a given mask, our method reduces misalignment and unnatural artifacts common in full-image generation.

2.3   Filtering uncertain synthetic regions

We propose an adaptive filtering mechanism to exclude uncertain synthetic regions. Specifically, uncertain synthetic regions are less likely to be correctly identified by a well-trained semantic segmentation model. Building on this observation, we filter synthetic pixels based on the following criterion: if a synthetic tumor pixel is not classified as a tumor pixel by a segmentation model pretrained on real dataset, it is marked as uncertain and excluded from the loss computation. The uncertain regions are defined by the false negative pixels defined as $FN = m - (m \cap P)$, where m is the ground-truth mask and P is the mask predicted by the pretrained segmentation model.

3   Experiment

3.1   Datasets

DCIS The first dataset is Ductal Carcinoma in Situ (DCIS) [22], a non-invasive form of breast cancer that can progress to invasive ductal carcinoma. This dataset includes 116 WSIs with annotations from 109 patients. We utilize images at $10\times$ magnification and split the slides into 81 for training, 17 for validation, and 18 for testing. To evaluate our method's performance on smaller datasets, we randomly sample 8 and 20 WSIs.

CATCH The second dataset is the Pan-tumor Canine Cutaneous Cancer Histology (CATCH) dataset [23]. It consists of 350 WSIs with annotations from 282 patients, with 50 slides per tumor type. We use images at $5\times$ magnification and adhere to the official dataset split, with 245 slides for training, 35 for validation, and 70 for testing. To assess our method's performance on reduced datasets, we randomly select 7 and 14 WSIs.

CAMELYON16 The final dataset is the CAMELYON16 challenge [5] dataset for breast cancer. We use images at $10\times$ magnification and follow the official data split, with 111 annotated WSIs for training and 48 for testing. From the training set, we set aside 11 slides for validation. To evaluate our method on smaller subsets, we randomly select 10 and 20 WSIs.

3.2   Implementation details

Our implementation builds on publicly available latent diffusion code for histopathology, using masks to control tumor inpainting and incorporating regional tumor embeddings from a different image as content conditioning via an SSP model [4]. Both training and testing are performed on non-overlapping patches extracted from each slide.
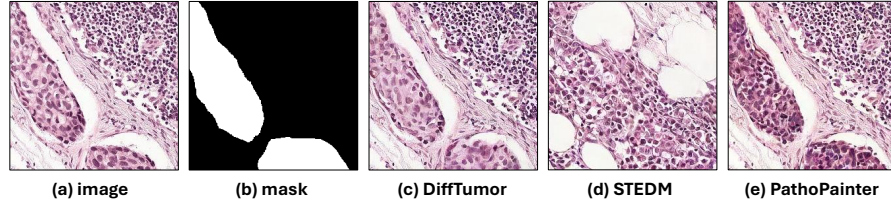
(a) image   (b) mask   (c) DiffTumor   (d) STEDM   (e) PathoPainter

Fig. 3: Visualization of generated images corresponding to the methods in Figure 1. The columns represent: (a) the original image x, (b) the ground truth mask m, and synthetic images generated by (c) DiffTumor [3], (d) STEDM [14], and (e) PathoPainter.

Latent Diffusion Model (LDM) In all experiments, we employ the LDM on the patches of 256×256 pixels. In particular, the pretrained VQ-VAE [18] with four downsampling blocks is used without any fine-tuning. For the U-Net denoiser, we use the model pretrained on ImageNet [19], as made available by [6], but fine-tuned it on our datasets for 20,000 training steps with a learning rate of 10^{-4}. To fine-tune the U-Net, we used two NVIDIA A100 GPUs and trained with a batch size of 100 per GPU.

Self-supervised pretrained (SSP) Model The SSP model employed in our work is HIPT [4], which is fine-tuned on our training sets. Specifically, we fine-tune the VIT_{256} model for 100 epochs using the AdamW [12] optimizer, with a batch size of 256 and a base learning rate of 0.0005. The first 10 epochs were used for warm-up to the base learning rate, followed by a decay phase governed by a cosine schedule. After fine-tuning, we extract features via mask pooling on the ViT feature map. We cluster all tumor embeddings from the training set into 10 groups using K-means [13].

Segmentation model For the downstream experiments, we utilize the widely used segmentation model, the Swin-Unet, configured with a Tiny architecture [2]. We perform binary segmentation to distinguish between foreground (tumor) and background (non-tumor). In experiments that involve training on synthetic images, we augment the real training set with synthetic images. The loss function consists of a weighted combination of cross-entropy loss and Dice loss, with both components assigned equal weights of 0.5. We apply traditional data augmentation, including flipping, rotation, and color adjustments for training. The models are trained on an NVIDIA TITAN Xp GPU for 10 epochs, with batch sizes of 12 for Swin-Unet. The model with the lowest validation loss is selected for testing.

3.3   Evaluation

To validate our synthetic images, we first performed a qualitative assessment, examining image-mask alignment, deviations from the original images, and biological plausibility. Following [25,3], we conducted a quantitative evaluation to

Table 1: Segmentation results across different datasets and slide numbers.

Dataset	Synthetic Data	IoU (%)	Variance	IoU (%)	Variance
		8 WSI		**20 WSI**	
DCIS	None	62.33	0.1543	64.85	0.1432
	DiffTumor [3]	62.43	0.1529	65.03	0.1430
	STEDM [14]	61.98	0.1541	64.88	0.1492
	PathoPainter	**63.67**	**0.1510**	**66.09**	**0.1412**
		7 WSI		**14 WSI**	
CATCH	None	81.88	0.1017	84.77	0.0888
	DiffTumor [3]	82.01	0.0984	84.92	0.0849
	STEDM [14]	82.23	0.1123	84.87	0.0852
	PathoPainter	**83.10**	**0.0941**	**85.82**	**0.0809**
		10 WSI		**20 WSI**	
CAMELYON16	None	70.96	0.1516	75.69	0.1348
	DiffTumor [3]	71.20	0.1542	76.21	0.1321
	STEDM [14]	71.08	0.1551	75.92	0.1338
	PathoPainter	**72.42**	**0.1501**	**77.69**	**0.1243**

verify the effectiveness of our method by assessing the performance of segmentation models trained on augmented datasets with synthetic images and tested on real data. For this evaluation, we used the Intersection over Union (IoU) score for the foreground class and analyzed the variance of IoU scores across samples.

Qualitative image generation results As shown in Figure 3, comparing our proposed PathoPainter with DiffTumor [3], the generated tumor region in (c) closely resembles the original image (a) in both staining and texture. However, in (e), the H&E stain appears darker, and the tumor texture is denser while still maintaining biological plausibility. When comparing PathoPainter with STEDM [14], the full image generated in (d) no longer aligns with the ground truth mask (b), indicating inconsistencies. In contrast, PathoPainter preserves accurate alignment with the mask (b), ensuring better structural consistency.

Histopathological images generated using PathoPainter effectively preserve the mask, ensuring the generation of precise, high-quality image-mask pairs. When integrated with real images, these synthetic samples enhance the augmented dataset, providing robust and well-aligned training data that significantly improve segmentation model performance. leveraging tumor embedding of a different image, PathoPainter introduces greater diversity into the dataset while maintaining biological plausibility, further enriching the training process.

Segmentation results We select DiffTumor [3] and STEDM [14] as representative state-of-the-art methods for segmentation data synthesis. Table 1 presents the performance of the Swin-Unet model across various datasets and training

Table 2: Analysis of the effects of different design choices on segmentation performance on DCIS [22] with the Swin-Unet. From left to right are the ablations on: (a) Comparison of embeddings from the SSP model using the full image vs. regional tumor. (b) Impact of embedding source: current image vs. other images. (c) Effectiveness of the filtering strategy for synthetic data.

SSP Embedding	IoU (%)	Embedding Source	IoU (%)	Filter	IoU (%)
Full image	61.28	Self	61.87	w/o.	61.62
Regional tumor	**63.67**	Other	**63.67**	w/.	**63.67**

data scales when incorporating synthetic training data from different models. Across all datasets (DCIS, CATCH, and CAMELYON16), incorporating our synthetic data consistently improves IoU scores compared to using real data alone (*e.g.*, 63.67% *vs.* 62.33% with 8 slides on DCIS). In contrast, STEDM results in a performance drop due to inaccurate layout generation. Similarly, TumorDiff shows only marginal improvement, as its tumor conditioning mechanism limits diversity. Furthermore, our method consistently outperforms other synthesis approaches, achieving the best overall performance across all datasets.

3.4   Ablation Studies

We conduct ablation studies on 8 slides of DCIS dataset.

Effectiveness of regional tumor embedding. Table 2(a) evaluates the impact of regional tumor embeddings, extracted using the SSP model and mask, on segmentation performance. Conditioning on regional tumor embeddings improves the IoU from 61.28% to 63.67% compared to full-image embeddings, demonstrating its effectiveness in capturing more precise semantic features directly relevant to tumor regions.

Effectiveness of embedding source. Table 2(b) shows the impact of embedding sources on model performance. Utilizing embeddings from a different image demonstrates clear advantages by introducing diversity into the training data.

Effectiveness of filter strategy. Table 2(c) demonstrates the effect of the filtering strategy, explained in Section 2.3, on segmentation performance. Applying the proposed filtering strategy results in a substantial improvement in IoU from 61.62% to 63.67%. This demonstrates that adaptively filtering out uncertain synthetic regions effectively enhances the quality of the training data, thereby improving segmentation accuracy and ensuring more consistent results.

4   Conclusion

In this work, we proposed PathoPainter, a novel framework for generating synthetic histopathology image-mask pairs to address the challenges of limited annotated datasets in tumor segmentation. By preserving the background and

conditioning image generation on tumor region, our method ensures alignment between the image and its corresponding mask. Additionally, we introduced a tumor embedding sampling mechanism to enlarge dataset diversity while preserving biological plausibility. To further improve data quality, we developed a simple yet effective filtering strategy to suppress uncertain synthesized regions. Comprehensive evaluations across multiple datasets demonstrated the effectiveness of our approach in enriching dataset diversity and enhancing segmentation performance, highlighting its potential to advance pathology image analysis.

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

References

1. Aversa, M., Nobis, G., Hägele, M., Standvoss, K., Chirica, M., Murray-Smith, R., Alaa, A.M., Ruff, L., Ivanova, D., Samek, W., Klauschen, F., Sanguinetti, B., Oala, L.: Diffinfinite: Large mask-image synthesis via parallel random patch diffusion in histopathology. In: NeurIPS. vol. 36, pp. 78126–78141 (2023)
2. Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M.: Swin-unet: Unet-like pure transformer for medical image segmentation. In: ECCV Workshops. pp. 205–218 (2022)
3. Chen, Q., Chen, X., Song, H., Xiong, Z., Yuille, A., Wei, C., Zhou, Z.: Towards generalizable tumor synthesis. In: CVPR. pp. 11147–11158 (2024). https://doi.org/10.1109/CVPR52733.2024.01060
4. Chen, R.J., Chen, C., Li, Y., Chen, T.Y., Trister, A.D., Krishnan, R.G., Mahmood, F.: Scaling vision transformers to gigapixel images via hierarchical self-supervised learning. In: CVPR. pp. 16123–16134 (2022). https://doi.org/10.1109/CVPR52688.2022.01567
5. Ehteshami Bejnordi, B., Veta, M., Johannes van Diest, P., van Ginneken, B., Karssemeijer, N., Litjens, G., van der Laak, J.A.W.M., the CAMELYON16 Consortium: Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. JAMA **318**(22), 2199–2210 (12 2017). https://doi.org/10.1001/jama.2017.14585
6. Graikos, A., Yellapragada, S., Le, M.Q., Kapse, S., Prasanna, P., Saltz, J., Samaras, D.: Learned representation-guided diffusion models for large-image generation. In: CVPR. pp. 8532–8542 (2024). https://doi.org/10.1109/CVPR52733.2024.00815
7. Han, C., Lin, J., Mai, J., Wang, Y., Zhang, Q., Zhao, B., Chen, X., Pan, X., Shi, Z., Xu, Z., Yao, S., Yan, L., Lin, H., Huang, X., Liang, C., Han, G., Liu, Z.: Multi-layer pseudo-supervision for histopathology tissue semantic segmentation using patch-level classification labels. Medical Image Analysis **80**, 102487 (2022). https://doi.org/https://doi.org/10.1016/j.media.2022.102487
8. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn (2018), https://arxiv.org/abs/1703.06870
9. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: NeurIPS. NIPS '20 (2020)
10. Kingma, D.P., Welling, M.: Auto-encoding variational bayes (2022), https://arxiv.org/abs/1312.6114

11. Liu, H., Yang, H., van Diest, P.J., Pluim, J.P.W., Veta, M.: WSI-SAM: Multi-resolution segment anything model (SAM) for histopathology whole-slide images (2024), https://arxiv.org/abs/2403.09257
12. Loshchilov, I.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)
13. MacQueen, J.: Some methods for classification and analysis of multivariate observations. In: Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability. vol. 1, pp. 281–297. Oakland, CA, USA (1967)
14. Öttl, M., Wilm, F., Steenpass, J., Qiu, J., Rübner, M., Hartmann, A., Beckmann, M., Fasching, P., Maier, A., Erber, R., Kainz, B., Breininger, K.: Style-extracting diffusion models for semi-supervised histopathology segmentation. In: Leonardis, A., Ricci, E., Roth, S., Russakovsky, O., Sattler, T., Varol, G. (eds.) ECCV. pp. 236–252 (2024)
15. Qaiser, T., Tsang, Y.W., Taniyama, D., Sakamoto, N., Nakane, K., Epstein, D., Rajpoot, N.: Fast and accurate tumor segmentation of histology images using persistent homology and deep convolutional features. Medical Image Analysis **55**, 1–14 (2019). https://doi.org/https://doi.org/10.1016/j.media.2019.03.014
16. Razavi, A., Van den Oord, A., Vinyals, O.: Generating diverse high-fidelity images with VQ-VAE-2. In: Advances in Neural Information Processing Systems (NeurIPS) (2019)
17. Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H.R., Albarqouni, S., Bakas, S., Galtier, M.N., Landman, B.A., Maier-Hein, K., et al.: The future of digital health with federated learning. NPJ Digital Medicine **3**(1), 119 (2020)
18. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: CVPR. pp. 10674–10685 (2022). https://doi.org/10.1109/CVPR52688.2022.01042
19. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A.C., Fei-Fei, L.: Imagenet large scale visual recognition challenge. International Journal of Computer Vision **115**(3), 211–252 (12 2015). https://doi.org/10.1007/s11263-015-0816-y
20. Schwarz, C.G., Kremers, W.K., Therneau, T.M., Sharp, R.R., Gunter, J.L., Vemuri, P., Arani, A., Spychalla, A.J., Kantarci, K., Knopman, D.S., et al.: Identification of anonymous mri research participants with face-recognition software. New England Journal of Medicine **381**(17), 1684–1686 (2019)
21. van Rijthoven, M., Balkenhol, M., Siliņa, K., van der Laak, J., Ciompi, F.: Hooknet: Multi-resolution convolutional neural networks for semantic segmentation in histopathology whole-slide images. Medical Image Analysis **68**, 101890 (2021). https://doi.org/https://doi.org/10.1016/j.media.2020.101890
22. Wetstein, S.C., Stathonikos, N., Pluim, J.P., Heng, Y.J., ter Hoeve, N.D., Vreuls, C.P., van Diest, P.J., Veta, M.: Deep learning-based grading of ductal carcinoma in situ in breast histopathology images. Laboratory Investigation **101**(4), 525–533 (2021). https://doi.org/https://doi.org/10.1038/s41374-021-00540-6
23. Wilm, F., Fragoso, M., Marzahl, C., Qiu, J., Puget, C., Diehl, L., Bertram, C.A., Klopfleisch, R., Maier, A., Breininger, K., Aubreville, M.: Pan-tumor CAnine cuTaneous Cancer Histology (CATCH) dataset. Scientific Data **9**(1), 588 (2022). https://doi.org/10.1038/s41597-022-01692-w
24. Xu, G., Song, Z., Sun, Z., Ku, C., Yang, Z., Liu, C., Wang, S., Ma, J., Xu, W.: Camel: A weakly supervised learning framework for histopathology image segmentation. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV). pp. 10681–10690 (2019). https://doi.org/10.1109/ICCV.2019.01078

25. Yang, L., Xu, X., Kang, B., Shi, Y., Zhao, H.: Freemask: synthetic images with dense annotations make stronger segmentation models. In: NeurIPS (2024)
26. Zhang, L., Wen, T., Min, J., Wang, J., Han, D., Shi, J.: Learning object placement by inpainting for compositional data augmentation. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIII 16. pp. 566–581. Springer (2020)