AVDM: Controllable Adversarial Diffusion Model for Vessel-to-Volume Synthesis

Jian Dai[1*], Wanchen Liu[2], Honghao Cui[1], Xiao Liu[2], Jiajun Wang[3], Zhiji Zheng[1] and Daoying Geng[1,2,4,5]

[1] College of Biomedical Engineering, Fudan University, Shanghai, China
[2] Department of Radiology, Huashan Hospital, Fudan University, Shanghai, China
[3] Department of Radiotherapy, Huashan Hospital, Fudan University, Shanghai, China
[4]Shanghai Engineering Research Center of Intelligent Imaging for Critical Brain Diseases, Shanghai, China
[5]Institute of Functional and Molecular Medical Imaging, Shanghai, China

jdai22@m.fudan.edu.cn

Abstract. 3D blood vessel segmentation remains a critical yet challenging task in medical image analysis. The heterogeneity of clinical imaging protocols introduces substantial domain gaps, limiting the generalizability of supervised learning methods that rely on manually annotated pixel-level labels for individual datasets. Furthermore, the large labeled volumetric datasets are difficult to collect because of data privacy issues. While diffusion models offer potential solutions by generating shareable synthetic data, existing approaches often exhibit poor alignment between synthesized volumes and their corresponding vascular structure input. To address these limitations, we propose **Controllable Adversarial Diffusion Model (AVDM)**, which integrates adversarial supervision into the diffusion training framework. Unlike conventional methods that generate imperceptible perturbations, AVDM synthesizes adversarial instances emphasizing structural variations critical for volume synthesis. Specifically, we design a segmentation-guided discriminator that enforces both the photorealism of generated volumes and pixel-level consistency with original vessel annotations. This supervision mechanism enables high-resolution synthesis of anatomically plausible vascular structures. Experiments demonstrate that AVDM surpasses state-of-the-art methods in generative fidelity and enhances performance on downstream tasks. Our code is available at https://github.com/jdai22/AVDM.

Keywords: diffusion models, image generation, vessel segmentation.

1 Introduction

Blood vessel segmentation is a vital task in medical image analysis, particularly for vascular disorders like stroke [1], cerebral aneurysms [2], and coronary disease [3], where it plays a crucial role in diagnosis and treatment. Despite advances in medical image analysis, accurate and robust segmentation of fully-connected vasculature in

task-specific imaging modalities remains a challenging problem. This is primarily due to the complexity introduced by intricate minuscule vascular geometries, as well as significant domain gaps caused by imaging modality and protocol-specific variations in signal-to-noise ratios, vascular pattern, and background tissues. These variations severely restrict the ability of supervised learning methods to generalize unseen 3D blood vessel domains [4]. Consequently, researchers and clinicians rely on the labor-intensive process of manually annotating pixel-level consistent labels from scratch to analyze vascular images.

Recently, generative models such as Generative Adversarial Networks (GANs) [5] and Denoising Diffusion Probabilistic Models (DDPMs) [6] have been widely applied to synthesize medical images, primarily due to the limited availability of real images. Generating an image from a mask is a type of image-to-image translation work [7]. However, Existing standard models fail to directly generate precise 3D volumes with corresponding vessel masks. When adopting powerful large pre-trained latent diffusion models (LDMs) [8] for vessel-to-volume synthesis, the fine-tuning model exhibits a domain shift from real-world noise due to the loss of mask controllability without considering the characteristics and heterogeneity geometric structure of the blood vessel. Consequently, the model can only generate samples similar or duplicates of the existing training set, thus adversely impacting its utility for potential downstream tasks requiring diverse data. Since annotated data only partially reflect real-world environments, synthetic samples are designed to complement real data by providing additional diversity.

Adversarial examples [9] mitigate the noise prediction errors due to poor alignment with condition input. In this context, we propose Controllable Adversarial Diffusion Modeling (AVDM), which introduces constrained adversarial supervision into the diffusion training process. We map images to a low-dimensional latent manifold [10] and shift them along gradient-optimized directions to generate adversarial examples with high structural fidelity and domain adaptability. To preserve texture and structural consistency with the original vessel during back-mapping, we use a semantic segmentation-based discriminator [11] that leverages conditional information for pixel-wise feedback to the diffusion model [12]. This integration ensures that the adversarial examples are realistically aligned and accurately correspond to their original mask labels.

We present AVDM, a diffusion-based framework for generating high-fidelity volumetric images conditioned on vessel mask inputs. This model represents a significant advancement in mask-conditional medical image generation, achieving anatomical precision while preserving realism. We evaluate AVDM performance across various datasets, demonstrating that it surpasses state-of-the-art mask-conditional generative models in fidelity to input anatomical masks.

2 Method

2.1 Projecting Volumetric to Diffusion Latent

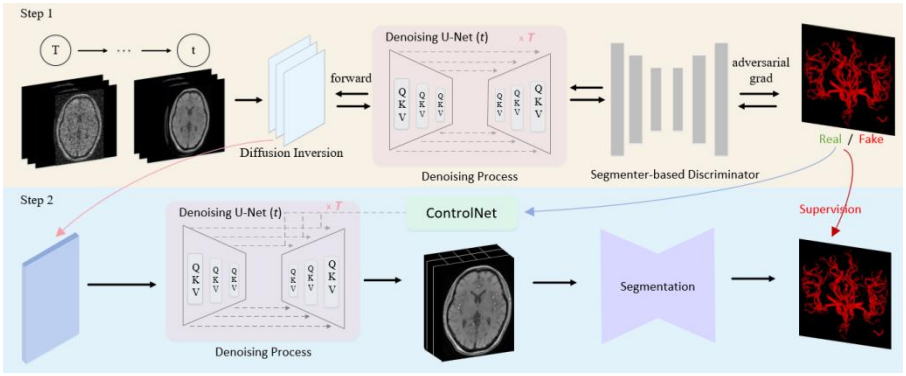

Fig. 1. The architecture of our proposed AVDM framework. In the first step, we project the input volume into the latent space and then optimized the latent space via adversarial discriminator supervision. In the second step, we use the optimized latent to generate adversarial samples controlled by input vessel masks.

To generate adversarial examples capable of addressing domain shifts across diverse domains, we leverage generative models like Stable Diffusion to project volumetric images into a low-dimensional manifold. By optimizing within this manifold space, we efficiently identify adversarial latent representations that are reprojected into the image space to synthesize consistent and diverse adversarial samples. Given an input volumetric image x_0, we employ diffusion inversion to map it to low-dimension latent space. The inversion utilizes a schedule $\{\beta_1, ..., \beta_T\} \in (0,1)$, within $\alpha_t = \prod_{i=1}^{t}(1 - \beta_i)$, where t represents the time steps. Specifically, we follow a forward diffusion procedure as follows:

$$x_{t+1} = \sqrt{\frac{\alpha_{t+1}}{\alpha_t}} x_t + \left(\sqrt{\frac{1}{\alpha_{t+1}} - 1} - \sqrt{\frac{1}{\alpha_t - 1}}\right) \cdot \epsilon_\theta(x_t, t, C) \quad (1)$$

Our framework operates along the reverse direction of the denoising process (i.e., $x_0 \rightarrow x_T$ rather than $x_T \rightarrow x_0$). Instead of progressively denoising from random noise to a clean image (forward process $q(X_t|X_{t-1})$), we project the input image x_0 into the latent space at a specific time point x_T via diffusion inversion.

The classifier-free guidance method [13] generates unconditional predictions and seamlessly merge them with predictions that are conditioned on specific inputs. Giving the guidance scale factor ω and null text embedding $\emptyset$, the classifier-free guided noise prediction at timestamp t is computed as:

$$\widetilde{\epsilon_\theta}(x_t, t, C, \emptyset) = \omega \cdot \epsilon_\theta(x_t, t, C) + (1 - \omega) \cdot \epsilon_\theta(x_t, t, \emptyset) \quad (2)$$

$\omega = 7.5$ is adopted as the standard setting for Stable Diffusion. In the reverse process of DDIM sampling, classifier-free guidance causing deviation from the learned noise distribution and generates visual anomalies that can diminish the realism of the output. To mitigate these limitations, we learn time-dependent null embeddings $\emptyset_{t=1}^{T}$ to preserve both distributional consistency and anatomical fidelity. Initially, executing the DDIM inverse sampling process with $\omega = 1$ yields a series of successive latent representations $\{x_0^*, \ldots, x_T^*\}$, starting with $x_0^* = x_0$. Subsequently we embark on an optimization process for the timesteps $t = \{T, \ldots, 1\}$,employing $\omega = 7.5$ and setting $\bar{x}_T = x_T^*$:

$$\min_{\emptyset_t} ||x_{t-1}^* - x_{t-1}(\bar{x}_t, t, C, \emptyset_t)||_2^2 \tag{3}$$

For ease of understanding, let $x_t - 1(\bar{x}_t, t, C, \emptyset_t)$ denote the DDIM sampling step, where $\bar{x}_t$ serves as the input latent, $\emptyset_t$ as the null text embedding, and C is the text embedding. Upon finishing each step, $\bar{x}_{t-1}$ is updated in accordance with the equation:

$$\bar{x}_{t-1} = x_{t-1}(\bar{x}_t, t, C, \emptyset_t) \tag{4}$$

Finally, we can achieve the latent representation $\bar{x}_T = x_T^*$ with the optimized null text embedding $\emptyset_t$ generated by the diffusion model. We exploit this latent in the low dimensional manifold to generate adversarial images.

2.2 Discriminator Supervision with Segmentation Alignment Process

Build upon the latent representation, we formalize the denoising process of our diffusion model as follow, A U-Net denoise ϵ_θ is trained to estimate the additive noise through mean squared error (MSE) objective:

$$\mathcal{L}_{noise} = \mathbb{E}_{\epsilon \sim N(0,I),y,t}[\|\epsilon - \epsilon_\theta(x_t, y, t)\|^2] = \mathbb{E}_{\epsilon,x_0,y,t}\left[\|\epsilon - \epsilon_\theta\left(\sqrt{\alpha_t}x_0 + \sqrt{1-\alpha_t}\epsilon, y\right)\|^2\right] \tag{5}$$

Besides the noisy image x_t and the time step t, the UNet takes the vessel input y as an additional input. Since y contains the blood vessel information derived from the original image x_0, it simplify the noise estimation and then implicitly guides the image synthesis during the denoising step. From x_t and the noise prediction ϵ_θ,we can generate a denoised version of the clean image $\hat{x}_0^{(t)}$ as:

$$\hat{x}_0^{(t)} = \frac{x_t - \sqrt{1-\alpha_t}\epsilon_\theta(x_t, y, t)}{\sqrt{\alpha_t}} \tag{6}$$

However, the absence of explicit supervision for layout fidelity in the training objective $\mathcal{L}_{\text{noise}}$ often leads to misalignment between generative images $\hat{x}_0^{(t)}$ and structural conditions y. Thus we seek direct supervision on $\hat{x}_0^{(t)}$ to enforce the layout alignment. To encourage diversity in addition to alignment, we make the segmenter trainable along with the UNet training. Inspired by, we formulate an adversarial game between the UNet and the segmenter. Specifically, the segmenter acts as a discriminator that is trained to classify per-pixel class labels of real images, using the

paired ground-truth label maps. Meanwhile, the discriminator classifies the fake images generated by the UNet as an additional 'fake' class. Since the primary task of the discriminator is to perform semantic segmentation, its training objective is based on the standard cross-entropy loss

$$L_{Dis} = -\mathbb{E}\left[\sum_{c=1}^{N}\gamma_c\sum_{i,j}^{II\times W} y_{i,j,c}\, log\left(Dis(x_0)_{i,jc}\right)\right] - \mathbb{E}\left[\sum_{i,j}^{II\times W} log\left(Dis\left(\hat{x}_0^{(t)}\right)_{i,j,c=N+1}\right)\right] \quad (7)$$

2.3 Image Generation with Controllable Adversarial

After deriving an adversarial latent representation, a reverse diffusion process can be applied to generate the final adversarial examples. We gain enhanced spatial shaping capabilities by incorporating ControlNet [14] into this reverse process. ControlNet improves the precision of task-specific conditioning within the denoising U-Net architecture of the Stable Diffusion model. The Stable Diffusion model is built around a U-Net framework, consisting of an encoder, a middle block, and a decoder, with each segment comprising 12 blocks. ControlNet augments this structure by creating a trainable copy of the 12 encoder blocks and the middle block from the original model. These blocks are distributed across four resolution levels, with three blocks per level. The outputs of these trainable blocks are then seamlessly integrated into the 12 skip connections and the middle block of the diffusion U-Net. This integration significantly enhances the model's ability to refine and tailor image characteristics with greater accuracy.

Table 1. Quantitative results on synthetic data.

Method	ADAM		CoW		IXI-HH	
	FID↓	MMD↓	FID↓	MMD↓	FID↓	MMD↓
HA-GAN	23.79	0.29	17.01	0.20	18.16	0.24
WGAN	22.03	0.26	16.08	0.18	17.40	0.20
WDM	17.75	0.22	15.58	0.16	15.17	0.17
MAISI	15.38	0.17	13.85	0.13	14.56	0.14
Ours	**13.95**	**0.14**	**10.63**	**0.11**	**12.26**	**0.09**

3 Experiments

We conducted experiments on publicly available 3D blood vessel datasets to synthesize data. The dataset consists of 127 TOF-MRA volumes collected from three distinct data centers: ADAM [15], CoW [16], and IXI-HH [17]. These datasets exhibit substantial diversity, particularly regarding acquisition devices, scanning protocols, magnetic field strengths, and spatial resolutions. Additionally, we evaluated zero-shot, one-shot, and few-shot segmentation tasks as downstream applications, utilizing three unseen 3D blood vessel datasets: IXI-Guys[17], ICBM [18] and LocH [19]. In this

context, 3D patches of size 128^3 were extracted from the whole brain scans. We employ a vessel-density-driven approach, where patches are selected based on the presence of vascular structures identified in ground-truth vessel masks. A sliding window algorithm scans the 3D volume, prioritizing regions with high vessel density (computed as the proportion of vessel voxels within a patch). For each patient, we select the 50 patches ranked by vessel density, ensuring at least 20% vessel coverage per patch to balance representation and diversity. For the adversarial example generation process, we configured the DDIM (Denoising Diffusion Implicit Models) with T=100 steps.

Table 2. Quantitative results for 3D blood vessel segmentation on three tasks: zero, one and few-shot.

Task	Method	IXI-Guys		ICBM		LocH	
		Dice↑	clDice	Dice↑	clDice↑	Dice↑	clDice↑
Zero-shot	HA-GAN	46.63	43.32	39.33	37.88	26.60	24.65
	WGAN	48.76	46.89	42.28	40.68	27.07	25.64
	WDM	53.15	51.87	46.11	43.79	32.53	30.07
	MAISI	56.33	54.15	50.83	48.29	35.66	34.10
	Ours	58.15	56.72	52.11	51.45	44.95	40.91
One-shot	HA-GAN	65.69	64.56	57.78	55.24	46.93	44.45
	WGAN	67.60	63.05	59.78	57.83	47.13	45.59
	WDM	70.27	68.06	62.16	58.36	51.72	48.29
	MAISI	72.11	70.89	65.96	63.43	54.60	52.08
	Ours	74.84	73.16	69.18	67.99	57.70	56.19
Few-shot	HA-GAN	73.04	71.08	68.15	66.90	63.90	61.78
	WGAN	75.80	73.57	69.83	67.28	65.16	63.73
	WDM	79.97	76.91	72.55	69.74	68.15	66.53
	MAISI	81.93	78.44	74.09	73.33	72.89	71.19
	Ours	**83.86**	**81.47**	**76.57**	**75.38**	**74.60**	**73.60**

3.1 Baseline & Evaluation

To validate the effectiveness of our proposed method, we compared it against several generative approaches serving benchmarks. HA-GAN [20], a hierarchical amortized GAN, simultaneously generates high-resolution volumes to mitigate memory constraints. WGAN [21] generates high-resolution volumes along with labels in an end-to-end paradigm. WDM [22] is a diffusion-based framework for medical image synthesis that leverages wavelet-decomposed images. MAISI [23] employs a 3D U-Net trained in the latent space of a 3D VAE. We evaluated performance using the

Fréchet Inception Distance (FID) and Maximum Mean Discrepancy (MMD) to measure the similarity between the distributions of real and generated images. Additionally, for downstream blood vessel segmentation tasks, we employed the Dice coefficient (Dice) to assess segmentation accuracy and the topology-aware centerline Dice (clDice) to evaluate the preservation of tubular structure and vascular connectivity.

3.2 Quantitative and Qualitative Results

Quantitative comparisons are summarized in Table 1. We evaluated generative quality using the FID for feature-level realism and MMD for distribution similarity. Our method achieves state-of-the-art performance with FID and MMD scores of 12.28 and 0.24, respectively, representing a 2.67 reducing in FID (from 14.95 to 12.28) and 0.03 reducing (from 0.14 to 0.11) in MMD compared to the second-best method MAISI [23]. This demonstrates that our adversarial samples are more realistic reconstructed from a well-optimized gradient guided by the adversarial supervision and highlight a significant improvement in generative fidelity, as shown in Figure 2. Furthermore, conventional approaches like HA-GAN [20] and WGAN [21] show degraded performance across all metrics (FID > 16.16, MMD > 0.18), highlighting their inherent limitations in modeling the complex intensity distributions and geometric constraints of medical volumetric data.

As shown in Table 2, we systematically asses the downstream segmentation performance of our synthesis approach across three distinct tasks: zero-shot, one-shot, and few-shot segmentation. Across all experiments, we adopt the nn-Unet [24] framework as a segmentation backbone to ensure standardized comparisons. Our method demonstrates exceptional zero-shot generalizability, achieving average Dice and clDice scores of 51.7% and 49.6% across three diverse unseen domains. Notably, our method outperforms MAISI [23] by 4.1 Dice and 3.9 clDice points, highlighting the effectiveness of our vessel morphology-aware inductive bias derived from high-fidelity synthetic training data. When adapted to one-shot and few-shot settings through fine-tuning, the proposed method further improves segmentation accuracy by 15.54 and 26.64 Dice points, respectively compared to its zero-shot baseline. This demonstrates the critical role of our synthesized vascular patterns in enhancing downstream tasks.

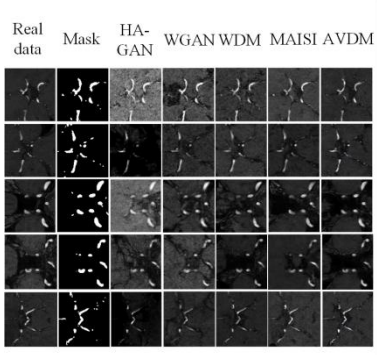

Fig. 2. Qualitative results on volume generation corresponding 3D vessel.

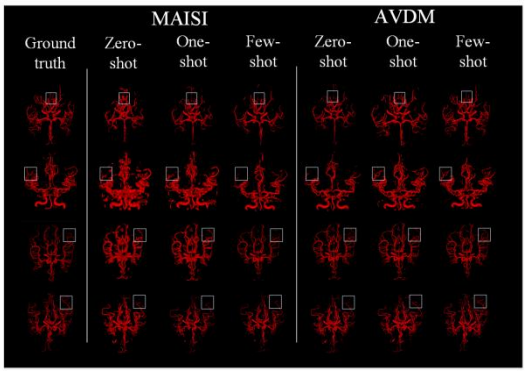

Fig. 3. Qualitative results on 3D vessel segmentation.

3.3 Ablation Studies

Main components. Table 3 show that performance significantly declines when latent projection is employed without Latent Optimization. This is primarily due to the lack of supervision, which prevents the projection from aligning with the model's learned patterns. Additionally, using only latent projection followed by reconstruction with ControlNet results in suboptimal performance, as this approach fails to incorporate model-specific insights in the projection process. In contrast, combining Latent Optimization and ControalNet achieves the best segmentation results.

Discriminator Ablation. We performed an ablation study to evaluate different discriminator designs, with results presented in Table 4. We explored two options: a CNN-based network [25] and a transformer-based network [26] — both of which enhance the fidelity of the baseline model. Additionally, rather than applying the discriminator solely in pixel space, we investigated a feature-space discriminator, which also demonstrated satisfactory performance.

Table 3. Ablation studies of main componen.

Latent Projection	Latent Optimization	Controllable Generation	Dice↑		
			IXI-Guys	ICBM	LocH
√			80.53	73.32	71.06
√	√		81.16	74.86	72.84
√		√	82.21	75.29	73.94
√	√	√	**83.86**	**76.57**	**74.60**

Table 4. Ablation on the discriminator type.

Method	FID↓		
	ADAM	CoW	IXI-HH
+ CNN-based UperNet	16.99	13.99	15.68
+ Transformer-based Segmenter	14.86	12.08	14.11
+ Feature-based	**13.95**	**10.63**	**12.26**

4 Conclusion

In this work, we propose AVDM, a novel framework that integrates adversarial supervision into a diffusion model to improve the faithfulness of vessel-to-volume synthesis. By leveraging a segmenter-based discriminator, we explicitly enforce morphological consistency with original vessel annotations through adversarial supervision, addressing the limitations of existing diffusion models that often fail to align synthesized volumes with vascular structure inputs. Our experiments demonstrates that ADVM achieves superior generative performance and enhance the downstream application.

Acknowledgments. This study was funded by the National Nature Science Foundation of China (82372048), the Science and Technology Commission of Shanghai Municipality (22TS1400900， 23S31904100， 22ZR1409500， 24SF1904200， 24SF1904201), the Qidong-Fudan Innovative Institute of Medical Sciences (KTB002).

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

References

1. Deshpande, A., Jamilpour, N., Jiang, B.: Automatic segmentation, feature extraction and comparison of healthy and stroke cerebral vasculature. NeuroImage (2020).
2. Nishi, H., Cancelliere, N. M., Rustici, A., et al.: Deep learning-based cerebral aneurysm segmentation and morphological analysis with three-dimensional rotational angiography. Journal of NeuroInterventional Surgery 16(2), 197-203 (2024)
3. Xian, Z., Wang, X., Yan, S., Yang, D., Chen, J., Peng, C.: Main Coronary Vessel Segmentation Using Deep Learning in Smart Medical. Mathematical Problems in Engineering 2020, 1–9 (2020)
4. Wittmann, B., Glandorf, L., Paetzold, J.C., Amiranashvili, T., W alchli, T., Razansky, D., Menze, B.: Simulation-Based Segmentation of Blood Vessels in Cerebral 3D OCTA Images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 645–655 (2024)
5. Esser, P., Rombach, R., Ommer, B.: Taming Transformers for High-Resolution Image Synthesis. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 1–12 (2021)
6. 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. 1–12 (2022)
7. Cheng, B., Liu, Z., Peng, Y., Lin, Y.: General image-to-image translation with one-shot image guidance. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 22736–22746 (2023)
8. Blattmann, A., Rombach, R., Ling, H., Dockhorn, T., Kim, S.W., Fidler, S., Kreis, K.: Align your latents: High-resolution video synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 22563–22575 (2023)
9. Dai, X., Liang, K., Xiao, B.: Advdiff: Generating unrestricted adversarial examples using diffusion models. In: European Conference on Computer Vision. pp. 93–109 (2024)
10. Li, X., Zhang, Z., Li, X., Chen, S., Zhu, Z., Wang, P., Qu, Q.: Understanding Diffusion-based Representation Learning via Low-Dimensional Modeling. In: NeurIPS 2024 Workshop on Mathematics of Modern Machine Learning. (2024)
11. Yang, M., Yang, R., Tao, S., Zhang, X., Wang, M.: Unsupervised Domain Adaptive Building Semantic Segmentation Network by Edge-Enhanced Contrastive Learning. Neural Networks 179, 106581 (2024)
12. Liang, Y., He, J., Li, G., Li, P., Klimovskiy, A., Carolan, N., Navalpakkam, V.: Rich human feedback for text-to-image generation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 19401–19411 (2024)

13. Shen, D., Song, G., Xue, Z., Wang, F.Y., Liu, Y.: Rethinking the spatial inconsistency in classifier-free diffusion guidance. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9370–9379 (2024)
14. 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)
15. Ham, S., Seo, J., Yun, J., Bae, Y. J., Kim, T., Sunwoo, L., Kim, N.: Automated Detection of Intracranial Aneurysms Using Skeleton-Based 3D Patches, Semantic Segmentation, and Auxiliary Classification for Overcoming Data Imbalance in Brain TOF-MRA. Scientific Reports 13(1), 12018 (2023).
16. Mou, L., Yan, Q., Lin, J., Zhao, Y., Liu, Y., Ma, S., Zhao, Y.: COSTA: A Multi-center TOF-MRA Dataset and A Style Self-Consistency Network for Cerebrovascular Segmentation. IEEE transactions on medical imaging (2024)
17. Bizjak, Ž., Chien, A., Burnik, I., Špiclin, Ž.: Novel Dataset and Evaluation of State-of-the-Art Vessel Segmentation Methods. Image Processing, 772–780. (2022)
18. Dumais, F., Caceres, M. P., Janelle, F., Seifeldine, K., Arès-Bruneau, N., Gutierrez, J., ... Whittingstall, K.: eICAB: A novel deep learning pipeline for Circle of Willis multiclass segmentation and analysis. Neuroimage p. 119425 (2022)
19. Rougé, P., Passat, N., Merveille, O.: Topology Aware Multitask Cascaded U-Net for Cerebrovascular Segmentation. PLoS ONE 19(12), e0311439 (2024)
20. Sun, L., Chen, J., Xu, Y., Gong, M., Yu, K., Batmanghelich, K.: Hierarchical Amortized GAN for 3D High Resolution Medical Image Synthesis. IEEE Journal of Biomedical and Health Informatics 26(8), 3966–3975 (2022)
21. Subramaniam, P., Kossen, T., Ritter, K., Hennemuth, A., Hildebrand, K., Hilbert, A., ... & Madai, V. I.: Generating 3D TOF-MRA Volumes and Segmentation Labels Using Generative Adversarial Networks. Medical Image Analysis p. 102396 (2022)
22. Friedrich, P., Wolleb, J., Bieder, F., Durrer, A., Cattin, P. C.: Wdm: 3d wavelet diffusion models for high-resolution medical image synthesis. In: MICCAI Workshop on Deep Generative Models. pp. 11–21 (2024)
23. Guo, P., Zhao, C., Yang, D., Xu, Z., Nath, V., Tang, Y., Xu, D.: Maisi: Medical ai for synthetic imaging. arXiv preprint arXiv:2409.11169 (2024)
24. Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J., Maier-Hein, K. H.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods **18**, 203–211 (2021)
25. Xiao, T., Liu, Y., Zhou, B., Jiang, Y., Sun, J.: Unified Perceptual Parsing for Scene Understanding. In: Computer Vision – ECCV 2018, Lecture Notes in Computer Science. pp. 432–448 (2018)
26. Strudel, R., Garcia, R., Laptev, I., Schmid, C.: Segmenter: Transformer for semantic segmentation. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 7262–7272 (2021)