

VDSB-GWSyn: Diffusion Schrödinger Bridge for Controllable and Anatomically Feasible Guidewire Synthesis in Coronary Angiography

Haoyuan Tang¹[0009-0002-3107-569X], Zhuo Zhang¹[0000-0002-3946-0720]*, Jialin Li¹[0009-0007-2608-1565], Shuai Xiao¹[0000-0003-4058-8120], and Jiachen Yang¹[0000-0003-2558-552X]

Tianjin University, Tianjin, China
z_zhuo@tju.edu.cn

Abstract. Coronary guidewire endpoint localization is a fundamental capability for computer-assisted PCI, and its importance increases as robot-assisted PCI is progressively adopted to reduce operator radiation exposure. However, the scarcity of annotated CAG images with guidewires and the limited adaptability of existing guidewire synthesis models remain key bottlenecks for guidewire endpoint localization. To address this issue, we propose VDSB-GWSyn, a Diffusion Schrödinger Bridge (DSB) model-based framework, enabling synthesis of controllable, high-fidelity guidewire samples under complex anatomical backgrounds. VDSB-GWSyn first uses our shape prior algorithm to learn the basic guidewire geometry. It then generates guidewire masks under constraints imposed by the vessel segmentation masks and outputs the corresponding endpoint coordinates. Finally, it synthesizes realistic guidewire samples on real CAG images using DSB conditioned with SPADE. Experimental results show that the guidewire samples synthesized by VDSB-GWSyn achieve favorable ROI-FID and ROI-KID, as well as high IPR scores. In addition, incorporating our synthesized data for synthetic pre-training followed by real fine-tuning substantially improves downstream guidewire endpoint localization, reducing MPE from 16.01 px to 7.71 px and increasing PCK at 3 px from 52.63% to 86.27%, leading to more clinically reliable deployment of robot-assisted guidewire delivery systems. Moreover, the core design philosophy of controllable device synthesis with strict background preservation and anatomical feasibility constraints has the potential to transfer to other interventional device perception tasks where annotated data are scarce.

Keywords: Coronary angiography · Guidewire synthesis · Diffusion Schrödinger Bridge

* Corresponding author.

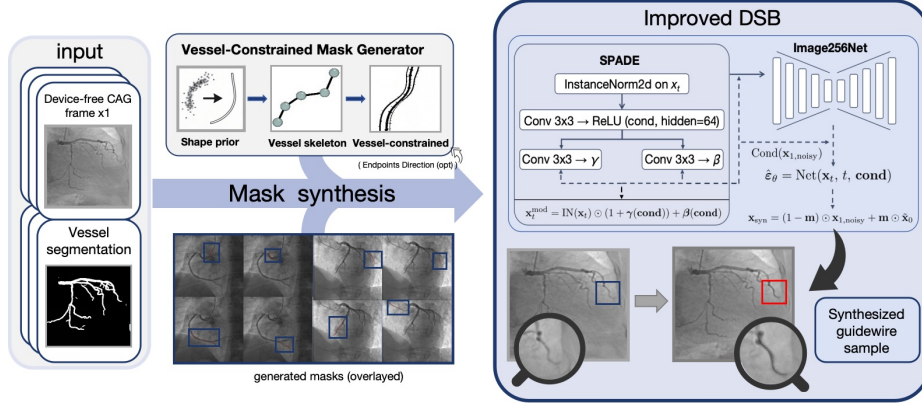


Fig. 1. Overview of the proposed VDSB-GWSyn framework for controllable guidewire synthesis.

1 Introduction

Reliable guidewire perception is a prerequisite for computer aided navigation in coronary angiography [14]. In fluoroscopy, the distal tip is often weakly visible and easily confounded by complex anatomical backgrounds [19], making end-point localization difficult and error prone [13]. Deep learning based endpoint localization requires large scale, high quality annotations, yet labeling guidewire pixels and tip coordinates is labor intensive and privacy constraints further limit data collection and sharing [5,28]. As a result, data scarcity remains a central bottleneck for robust endpoint localization in real world fluoroscopy.

Prior work has explored learning-based catheter and guidewire segmentation under limited annotations and augmentation, and has also investigated synthetic fluoroscopic data and transfer learning to reduce labeling requirements [1,7]. For image synthesis, unpaired adversarial translation has been applied to catheter synthesis in X-ray fluoroscopy to reduce pixel-level labeling effort, and adversarial augmentation has been used to improve catheter segmentation [22,23]. More recently, diffusion-based generators have demonstrated improved controllability through mask conditioning, and have been used to synthesize labeled fluoroscopy videos for guidewire segmentation [5,6,10,19]. Despite these advances, two challenges remain for guidewire endpoint localization. First, existing conditional synthesis pipelines typically assume that a valid guidewire mask is already available as input, while the anatomical feasibility of the guidewire trajectory within coronary vessels is not explicitly enforced inside the synthesis framework [19,23]. Second, despite recent progress in conditional synthesis, it remains challenging to achieve sufficient diversity in guidewire trajectories and adequate coverage of CAG background complexity, especially under limited annotation budgets. As a result, current augmentation pipelines may not sufficiently cover real CAG examples with complex vessel context and low guidewire visibility.

In addition, cost-efficient and scalable generation of guidewire CAG data with directly usable supervision, for example guidewire masks and endpoint coordinates, is still underexplored, limiting the effectiveness of synthetic augmentation for guidewire endpoint localization. Recent CAG-specific studies have further explored geometry-aware data synthesis for stenosis editing and auto-annotation generation for multi-view translational correction, highlighting the growing importance of controllable data generation in coronary angiography [3,12].

To address these gaps, we propose **VDSB-GWSyn**, a vessel-constrained diffusion Schrödinger bridge framework for controllable guidewire data synthesis on real coronary angiography backgrounds. Our method first learns guidewire shape priors and then synthesizes guidewire masks under vessel segmentation constraints, while directly producing endpoint coordinates. We formulate guidewire synthesis as a targeted inpainting problem so that generation is confined to the masked region and the anatomical background remains undisturbed. We further integrate SPADE to better preserve spatial conditioning information that can otherwise be weakened by normalization layers, improving local alignment between synthesized guidewire appearance and background texture [20]. Finally, we redesign the training objective with region normalization to mitigate gradient dilution caused by extreme guidewire pixel sparsity. Experiments demonstrate that pre-training with VDSB-GWSyn synthetic data greatly improves downstream endpoint localization on real-only evaluation sets.

2 Methods

The proposed VDSB-GWSyn framework is an end-to-end controllable synthesis pipeline designed for coronary angiography (CAG). As illustrated in Fig. 1, the framework consists of an Anatomical Mask Generator that samples plausible guidewire geometries and a Mask-Guided Diffusion Schrödinger Bridge (DSB) Engine that performs targeted rendering to ensure background consistency.

2.1 Controllable Mask Generation with Anatomical Constraints

To enable large scale synthesis of anatomically feasible guidewire masks under vessel constraints, we propose a geometry controllable mask generation algorithm.

Shape Prior and Radius Modeling. We learn a global shape prior from clinical masks. Let $\theta \in \mathbb{R}^{2N}$ be a vector representing N skeletonized control points. We extract a low-dimensional manifold $z = \text{PCA}(\theta)$ and model its distribution using a Gaussian Mixture Model (GMM). During sampling, the guidewire radius r is derived from a log-normal prior, $r = \exp(\mathcal{N}(\mu_r, \sigma_r^2))$, and capped by the local vessel diameter to adapt to distal branches.

Vessel-Constrained Path Sampling. Given a vessel mask ν , we construct an 8-connectivity graph G on its skeleton. We utilize Dijkstra’s algorithm to search for a base path P , where a penalty term is applied to junction nodes (degree ≥ 3) to suppress unrealistic bifurcations. The algorithm supports precise J-tip control

by allowing users to specify a tip location (x, y) and orientation ϕ . Sampled paths are projected into the vessel lumen using a distance-transform (DT) field. To simulate clinical diversity, we introduce a “wall-hugging” parameter $\rho \in [0, 1]$ that controls the target DT value during projection:

$$D_{\text{target}} = (1 - \rho)D_\nu + \rho d_{\min}, \quad (1)$$

where D_ν denotes the local vessel DT value and $d_{\min}$ is the minimum clearance to the vessel boundary. Thus, $\rho = 0$ keeps the guidewire closer to the vessel centerline, while larger ρ moves it toward the vessel wall without leaving the lumen. This module naturally supports multi-wire synthesis by iteratively sampling disjoint paths within the same anatomical domain.

2.2 Targeted Inpainting via DSB and SPADE Modulation

Standard diffusion models tend to redraw the entire image, which can cause texture drift in complex anatomical backgrounds. We therefore formulate guidewire synthesis as a targeted inpainting problem, using the original CAG anatomy as a hard background anchor [4,17].

Mask-Directed Noise Injection. Given a device-free background frame x_{bg} and a generated guidewire mask $m \in \{0, 1\}^{H \times W}$, we construct a noisy conditional endpoint x_1^n by confining the perturbation strictly within the mask region:

$$x_1^n = (1 - m) \odot x_{\text{bg}} + m \odot (x_{\text{bg}} + \sigma \epsilon), \quad \epsilon \sim \mathcal{N}(0, \mathbf{I}). \quad (2)$$

By keeping the $(1 - m)$ background region noise-free, the model treats the surrounding anatomy as a fixed reference. The diffusion Schrödinger bridge (DSB) [4,16] then learns an evolution path that renders guidewire appearance only within the mask m .

Spatial Feature Alignment. To achieve seamless integration between the synthesized guidewire and the original background, we incorporate Spatially-Adaptive Denormalization (SPADE) [20]. Specifically, for a normalized feature map $\hat{h}$ at an intermediate layer, SPADE modulates it with spatially varying parameters predicted from the conditioning signal:

$$\text{SPADE}(\hat{h} | x_1^n) = \gamma(x_1^n) \odot \hat{h} + \beta(x_1^n), \quad (3)$$

where $\gamma(\cdot)$ and $\beta(\cdot)$ are learnable functions producing per-location scaling and bias. This modulation encourages the diffusion network to align synthesized guidewire edges with local contrast and noise statistics of the specific CAG frame, while avoiding undesired redrawing of the underlying vascular structure.

2.3 Region-Normalized Objective for Sparse Structures

Guidewires are extremely sparse in CAG frames, often occupying $< 1\%$ of the pixels. To prevent the background from dominating the gradient, we redesign

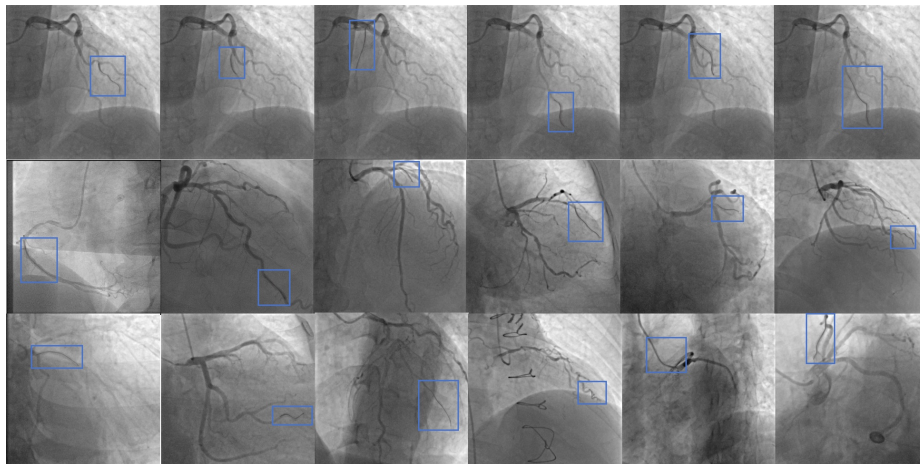


Fig. 2. Qualitative visualization of controllable guidewire synthesis in coronary angiography (CAG). **Top row:** Using the same device-free CAG background, we synthesize guidewires at different vessel locations to demonstrate spatial diversity while preserving the underlying anatomy. **Middle row:** Synthesized guidewires under different camera angles, showing robustness to view-dependent appearance variations in CAG. **Bottom row:** Synthesized guidewires across backgrounds with different levels of anatomical and noise complexity, illustrating background preservation and realism under challenging fluoroscopic conditions.

the training objective using region-normalization. The total loss is partitioned into instrument and background components:

$$\mathcal{L} = \lambda_w \mathcal{L}_{\text{wire}} + \lambda_b \mathcal{L}_{\text{bg}}, \quad (4)$$

where

$$\mathcal{L}_{\text{wire}} = \frac{1}{\sum_i m_i} \sum_i m_i \|\hat{y}_i - y_i\|_2^2. \quad (5)$$

By setting $\lambda_w \gg \lambda_b$, we force the model to focus on the high-frequency textures of the guidewire. Furthermore, we introduce a Masked Total Variation (TV) regularization term, acting only on the mask boundary, to enforce structural continuity and mitigate wire fragmentation in low-contrast regions.

3 Experiments

3.1 Datasets

Internal Real Dataset. We curate an internal coronary angiography (CAG) dataset consisting of 1,125 frames in total. Among them, 283 real frames containing clinical guidewires are manually annotated with pixel-level masks and

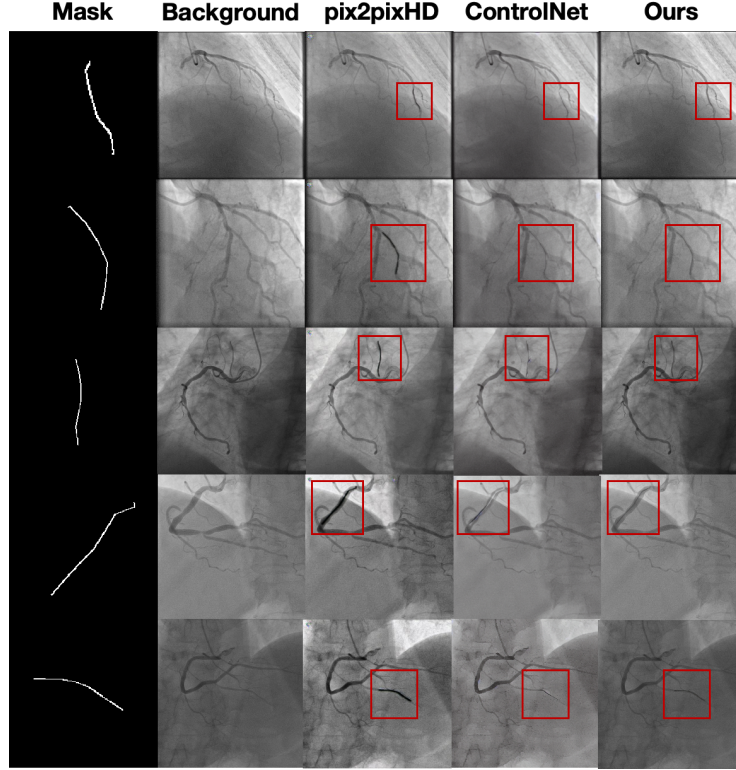


Fig. 3. Qualitative comparison of guidewire synthesis. Given the same device-free CAG background and the same input guidewire mask, we compare pix2pixHD, ControlNet, and our VDSB-GWSyn. Red boxes indicate the synthesized guidewire regions.

endpoint coordinates (tip and tail). This annotated subset serves three purposes: (i) extracting guidewire geometry for the shape-prior algorithm, (ii) providing paired samples to train the diffusion renderer, and (iii) acting as the training and evaluation set for the downstream endpoint localization task. For the downstream task, the 283 annotated frames are split into 182/50/51 for training/validation/testing.

Vessel Segmentation. The CAG images with vessel segmentation masks used for synthesis are drawn from two sources: (i) 955 internal clinical frames whose vessel masks are obtained using U-Mamba [18], and (ii) 832 high-quality frames selected from the ARCADE dataset [21] (released under the CC0-1.0 license), where vessel masks are generated using the same model and manually filtered for anatomical accuracy. The vessel masks are used exclusively to constrain guidewire mask sampling and to enforce anatomical plausibility; they are not used as inputs to the downstream endpoint localizer.

Table 1. Comprehensive comparison with baselines and ablation study on ROI-based generative quality metrics. Lower is better for ROI-FID / ROI-KID; higher is better for IPR (Precision/Recall). Bold indicates the best performance.

Method	ROI-KID (%) ↓	IPR (P/R) (%) ↑	ROI-FID ↓
<i>Baselines</i>			
pix2pixHD	5.5668 ± 0.1681	31.75 / 45.39	72.6365
ControlNet	2.1479 ± 0.1029	35.43 / 50.71	53.1129
<i>Ablation</i>			
w/o SPADE & w/o loss	2.7071 ± 0.1693	76.35 / 55.72	58.4721
w/o SPADE	2.5614 ± 0.1860	76.93 / 57.01	51.6956
Ours (DSB + SPADE)	1.6181 ± 0.1033	78.74/65.87	49.1371

Synthesis Training Set. Using 1,787 CAG frames paired with vessel segmentation masks, we first sampled 7,148 candidate guidewire masks (four per frame) and then retained 3,506 for rendering after filtering out anatomically implausible cases, where the sampled trajectories deviated from the vessel course or extended outside the vessel lumen due to segmentation inaccuracies. Then we constructed a 3000-image subset for downstream pre-training via manual quality control, excluding cases where the synthesized guidewire is missing, severely blurred, or strongly fragmented. This screening is applied uniformly and uses only image-level inspection with no downstream model feedback. The synthetic set covers diverse guidewire locations within the same CAG frame, varying clinical camera angles and different levels of background complexity. Representative examples are shown in Fig. 2. Following the same uniform image-level screening procedure, we also construct a 2,000-image subset for baseline synthetic pre-training to ensure a fair comparison.

3.2 Experimental Setup

All images are resized to 512×512 . We first fit a guidewire shape prior by skeletonizing clinical masks into 8 control points. The coordinates are compressed via PCA to a 4-dimensional latent space, and the latent distribution is modeled with a 5-component GMM. Our diffusion renderer follows a diffusion Schrödinger bridge with 1,000 discretization steps, noise levels from $t_0 = 10^{-4}$ to $T = 1.0$, and a symmetric beta set to 0.3. During training, Gaussian noise is injected only inside the guidewire mask regions to enforce localized inpainting. We optimize with AdamW ($\text{lr} = 5 \times 10^{-5}$), batch size 8, for 20k iterations, using EMA 0.99. We additionally apply a masked TV regularizer with weight 0.05 to preserve thin-structure continuity. For the downstream task, we adopt the RTMPose-S estimator with SimCC label encoding [9,15]. The model is pre-trained on three synthetic sets (1,000, 2,000, and 3,000 images) and then fine-tuned on real data.

Table 2. Downstream guidewire endpoint localization results of our VDSB-GWSyn. Lower is better for MPE; higher is better for PCK.

Pre-train data	MPE ↓	PCK@3 ↑	PCK@6 ↑	TIP MPE ↓	TAIL MPE ↓
0	16.0136	52.63	78.95	19.7241	12.3032
1,000	9.7175	77.45	84.31	9.6383	9.7966
2,000	7.7126	86.27	91.18	5.8726	9.5526
3,000	7.7781	85.29	88.24	6.7834	8.7727

3.3 Generation Evaluation

We evaluate the synthesis quality using both distribution alignment and local structure fidelity. Since coronary guidewires are extremely slender, global metrics such as standard FID [8] and KID [2] exhibit negligible variation. Following instance-level evaluation [24,26], we compute FID and KID on mask-centered ROI patches, which are extracted by cropping a tight bounding box around the synthesized guidewire mask, to focus evaluation on guidewire appearance rather than background. We also report IPR [11] to quantify manifold-based precision and recall, reflecting realism and coverage of the real ROI distribution. As shown in Table 1, our method achieves the best ROI-based performance compared to representative conditional generation baselines, including pix2pixHD [25] and ControlNet [29]. These quantitative trends are consistent with the qualitative comparisons in Fig. 3. Both baselines are prone to noticeable background drift and inconsistent local contrast, while pix2pixHD often produces overly thick wires with unnatural device-background blending, and ControlNet frequently yields fragmented or vanishing wire segments with weak structural continuity. In contrast, our method preserves background statistics while generating more coherent, anatomically plausible guidewire trajectories with fewer blending artifacts. Ablation results in Table 1 show that ROI-KID improves markedly when incorporating SPADE, indicating that spatially adaptive normalization enhances local feature alignment. Moreover, the proposed region-normalized objective mitigates background-dominated gradients, reducing wire fragmentation and improving structural continuity in synthesized samples.

3.4 Downstream Tasks

We adopt a synthetic pre-training followed by real fine-tuning protocol to evaluate the effectiveness of synthetic augmentation. RTMPose consistently benefits from the synthetic data. With the largest gain achieved at 2,000 synthesized images, it shows that MPE is reduced from 16.01 px to 7.71 px, and PCK at a 3-pixel threshold increases from 52.63% to 86.27% [27]. For a fair comparison, we apply the same protocol to the baseline methods. We first pre-train RTMPose on 2,000 synthetic images generated by pix2pixHD or ControlNet and then fine-tune the model on the same real training set using identical settings. Both baselines transfer poorly and yield large localization errors, with pix2pixHD

reaching an MPE of 34.71 px and PCK@3 px of 5.88% and ControlNet reaching an MPE of 41.68 px and PCK@3 px of 9.86%. This gap aligns with their typical synthesis artifacts, where pix2pixHD commonly alters the wire width and local contrast, whereas ControlNet often exhibits instability on slender regions, leading to incomplete wire traces in challenging frames.

4 Conclusion

We presented VDSB-GWSyn, a controllable guidewire synthesis framework for coronary angiography, which enables large-scale generation of vessel-constrained guidewire samples under complex coronary backgrounds. The synthesized guidewire samples generated by our model substantially improve endpoint localization on real CAG images. Despite these encouraging results, several limitations remain. In particular, the quality of automatically generated guidewire masks still depends on the accuracy of vessel segmentation. In cases of spatially overlapping vessels or imperfect lumen delineation, sampled trajectories may enter anatomically inconsistent branches, leading to implausible masks that require manual rejection. Our future work will focus on topology-aware constraints and uncertainty-aware sampling to improve robustness and scalability.

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

References

1. Ambrosini, P., Ruijters, D., Niessen, W.J., Moelker, A., van Walsum, T.: Fully automatic and real-time catheter segmentation in x-ray fluoroscopy. In: International conference on medical image computing and computer-assisted intervention. pp. 577–585. Springer (2017)
2. Bińkowski, M., Sutherland, D.J., Arbel, M., Gretton, A.: Demystifying mmd gans. arXiv preprint arXiv:1801.01401 (2018)
3. Cao, Y., Zhang, Z., Xiao, S., Li, J., Lan, G., Wen, J., Yang, J.: Automatic translational correction of multi-view coronary angiography based on auto-annotation data generation. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 40, pp. 2652–2660 (2026)
4. De Bortoli, V., Thornton, J., Heng, J., Doucet, A.: Diffusion schrödinger bridge with applications to score-based generative modeling. *Advances in neural information processing systems* **34**, 17695–17709 (2021)
5. Dorjsembe, Z., Pao, H.K., Odonchimed, S., Xiao, F.: Conditional diffusion models for semantic 3d brain mri synthesis. *IEEE Journal of Biomedical and Health Informatics* **28**(7), 4084–4093 (2024)
6. Du, Y., Jiang, Y., Tan, S., Wu, X., Dou, Q., Li, Z., Li, G., Wan, X.: Arsdm: colonoscopy images synthesis with adaptive refinement semantic diffusion models. In: International conference on medical image computing and computer-assisted intervention. pp. 339–349. Springer (2023)

7. Gherardini, M., Mazomenos, E., Mencias, A., Stoyanov, D.: Catheter segmentation in x-ray fluoroscopy using synthetic data and transfer learning with light u-nets. *Computer Methods and Programs in Biomedicine* **192**, 105420 (2020)
8. 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)
9. Jiang, T., Lu, P., Zhang, L., Ma, N., Han, R., Lyu, C., Li, Y., Chen, K.: Rtm-pose: Real-time multi-person pose estimation based on mmpose. *arXiv preprint arXiv:2303.07399* (2023)
10. Konz, N., Chen, Y., Dong, H., Mazurowski, M.A.: Anatomically-controllable medical image generation with segmentation-guided diffusion models. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 88–98. Springer (2024)
11. Kynkäänniemi, T., Karras, T., Laine, S., Lehtinen, J., Aila, T.: Improved precision and recall metric for assessing generative models. *Advances in neural information processing systems* **32** (2019)
12. Li, J., Zhang, Z., Cao, Y., Lan, G., Wen, J., Xiao, S., Yang, J.: Geometrically constrained stenosis editing in coronary angiography via entropic optimal transport. *arXiv preprint arXiv:2605.08851* (2026)
13. Li, R.Q., Xie, X.L., Zhou, X.H., Liu, S.Q., Ni, Z.L., Zhou, Y.J., Bian, G.B., Hou, Z.G.: Real-time multi-guidewire endpoint localization in fluoroscopy images. *IEEE Transactions on Medical Imaging* **40**(8), 2002–2014 (2021)
14. Li, R.Q., Xie, X.L., Zhou, X.H., Liu, S.Q., Ni, Z.L., Zhou, Y.J., Bian, G.B., Hou, Z.G.: A unified framework for multi-guidewire endpoint localization in fluoroscopy images. *IEEE Transactions on Biomedical Engineering* **69**(4), 1406–1416 (2021)
15. Li, Y., Yang, S., Liu, P., Zhang, S., Wang, Y., Wang, Z., Yang, W., Xia, S.T.: Simcc: A simple coordinate classification perspective for human pose estimation. In: *European conference on computer vision*. pp. 89–106. Springer (2022)
16. Liu, G.H., Vahdat, A., Huang, D.A., Theodorou, E.A., Nie, W., Anandkumar, A.: I²sb: Image-to-image schrödinger bridge. *arXiv preprint arXiv:2302.05872* (2023)
17. Lugmayr, A., Danelljan, M., Romero, A., Yu, F., Timofte, R., Van Gool, L.: Repaint: Inpainting using denoising diffusion probabilistic models. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. pp. 11461–11471 (2022)
18. Ma, J., Li, F., Wang, B.: U-mamba: Enhancing long-range dependency for biomedical image segmentation. *arXiv preprint arXiv:2401.04722* (2024)
19. Pan, S., Liu, Y., Zhao, L., Chen, E.Z., Chen, X., Chen, T., Sun, S.: Label-efficient data augmentation with video diffusion models for guidewire segmentation in cardiac fluoroscopy. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. vol. 39, pp. 6308–6316 (2025)
20. 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)
21. Popov, M., Amanturdieva, A., Zhaksylyk, N., Alkanov, A., Saniyazbekov, A., Aimyshev, T., Ismailov, E., Bulegenov, A., Kuzhukeyev, A., Kulanbayeva, A., et al.: Dataset for automatic region-based coronary artery disease diagnostics using x-ray angiography images. *Scientific data* **11**(1), 20 (2024)
22. Ullah, I., Chikontwe, P., Choi, H., Yoon, C.H., Park, S.H.: Synthesize and segment: Towards improved catheter segmentation via adversarial augmentation. *Applied Sciences* **11**(4), 1638 (2021)

23. Ullah, I., Chikontwe, P., Park, S.H.: Catheter synthesis in x-ray fluoroscopy with generative adversarial networks. In: International Workshop on PRedictive Intelligence In MEdicine. pp. 125–133. Springer (2019)
24. Wang, D., Zhang, Y., Bai, B., Yu, X., Shu, X., Dai, Y.: Grade: A generalization robustness assessment via distributional evaluation for remote sensing object detection. *Remote Sensing* **17**(22), 3771 (2025)
25. 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)
26. Xu, J., Wang, S., Chen, J., Li, Z., Jia, P., Zhao, F., Xiang, G., Hao, Z., Zhang, S., Xie, X.: Decouple distortion from perception: Region adaptive diffusion for extreme-low bitrate perception image compression. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 18051–18061 (2025)
27. Yang, Y., Ramanan, D.: Articulated human detection with flexible mixtures of parts. *IEEE transactions on pattern analysis and machine intelligence* **35**(12), 2878–2890 (2012)
28. Yazdani, M., Medghalchi, Y., Ashrafian, P., Hacihaliloglu, I., Shahriari, D.: Flow matching for medical image synthesis: Bridging the gap between speed and quality. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 216–226. Springer (2025)
29. 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)