CriDiff: Criss-cross Injection Diffusion Framework via Generative Pre-train for Prostate Segmentation

Tingwei Liu, Miao Zhang, Leiye Liu, Jialong Zhong, Shuyao Wang, Yongri Piao (✉), and Huchuan Lu

Dalian University of Technology, China
tingweiliu@mail.dlut.edu.cn, yrpiao@dlut.edu.cn

Abstract. Recently, the Diffusion Probabilistic Model (DPM)-based methods have achieved substantial success in the field of medical image segmentation. However, most of these methods fail to enable the diffusion model to learn edge features and non-edge features effectively and to inject them efficiently into the diffusion backbone. Additionally, the domain gap between the images features and the diffusion model features poses a great challenge to prostate segmentation. In this paper, we proposed CriDiff, a two-stage feature injecting framework with a Criss-cross Injection Strategy (CIS) and a Generative Pre-train (GP) approach for prostate segmentation. The CIS maximizes the use of multi-level features by efficiently harnessing the complementarity of high and low-level features. To effectively learn multi-level of edge features and non-edge features, we proposed two parallel conditioners in the CIS: the Boundary Enhance Conditioner (BEC) and the Core Enhance Conditioner (CEC), which discriminatively model the image edge regions and non-edge regions, respectively. Moreover, the GP approach eases the inconsistency between the images features and the diffusion model without adding additional parameters. Extensive experiments on four benchmark datasets demonstrate the effectiveness of the proposed method and achieve state-of-the-art performance on four evaluation metrics. The source code will be publicly available at https://github.com/LiuTingWed/CriDiff.

Keywords: Deep learning · Diffusion models · Prostate segmentation

1 Introduction

Prostate cancer, as the second most common cancer affecting men, necessitates accurate diagnostic tools for effective management [17]. Precise segmentation of the prostate is critical for the diagnosis and treatment planning of prostate cancer. With the development of deep learning, convolutional neural networks (CNNs) have made significant progress for prostate segmentation [7, 14, 18]. Although the above methods achieve promising results, they use the softmax in the cross-entropy loss overemphasizes the highest logit, leading to deterministic predictions. However, estimating the model's output uncertainty is crucial for

clinical doctors to further diagnose uncertain areas, because medical segmentation problems are often characterized by ambiguities and multiple hypotheses may be plausible [21].

Recently, Diffusion Probabilistic Models (DPMs) have led to unprecedented advancements in content generation tasks. Because they have the capability to generate different predictions by running multiple times, many DPM-based methods [2, 22–24] are proposed in the field of medical image segmentation. Despite these methods have shown great performances, the intricate anatomical positioning of the prostate, along with its visual similarity to adjacent tissues, presents significant challenges on accurate prediction of edge [26]. However, these methods overlook the learning of boundary information and treat all regions with equal importance. DermoSegDiff [2] introduces a novel boundary loss function by calculating the distance between each foreground pixel in the ground-truth label and the nearest background pixel. However, this weighted loss requires careful adjustment on the coefficients to balance the learning between edge and non-edge areas, relying on laborious trial and error. Moreover, previous DPM-based methods inject the multi-level features of medical images stage by stage into the diffusion backbone (*e.g.*, high-level semantic features are injected into deeper layers and low-level features are injected to shallower layers). This approach leads to the underutilization of multi-level features, limiting early-stage accuracy in object localization or shaping and impeding the model's capability to generate fine-grained objects in later stages. Therefore, it is essential to design a strategy that effectively learns multi-level features of edges and non-edges and enhances utilization of these features when integrating them into the diffusion model.

When the diffusion model applies to segmentation tasks, randomly initialized diffusion model parameters diffuse the final prediction map under the guidance of image conditional features from the specific data domain. The difference between diffusion model features and conditional features creates a domain gap, especially pronounced in prostate images. This domain gap impedes model convergence and diminishes performance. Current DPM-based approaches in medical image segmentation have not fully considered this issue. Consequently, it is essential to introduce an efficient method to reduce the domain gap, enabling better feature learning in diffusion models.

To address the aforementioned problems, we proposed a novel two-stage framework with a feature injection strategy and a generative pre-train method for prostate segmentation, entitled CirDiff. Specifically, we proposed the Criss-cross Injection Strategy (CIS) for enabling the diffusion to complementarily utilize multi-level features of edges and non-edge areas. To this goal, we proposed two parallel conditioners in this strategy, named Boundary Enhance Conditioner (BEC) and Core Enhance Conditioner (CEC). These two conditioners with distinct structures, are capable of discriminative learning of image edge features and other non-edge regions features. The BEC employs a triangular architecture that progressively cuts down the number of layers, focusing on learning edge textures features. Conversely, the CEC focuses on learning non-edge semantic features

via an inverted triangular architecture that progressively increases the number of layers. Then, we injected them into the diffusion backbone in a crisscross manner, promoting the diffusion model ability of learning edges and non-edge areas. Furthermore, we introduced a Generative Pre-train (GP) approach for prostate segmentation. The GP pretrains the diffusion model on generative tasks within the target domain, aligning feature representations more closely with the target domain. This approach narrows the domain gap between the conditional features and the diffusion model features, thus improving model performance without introducing additional parameters. In brief, the contributions of this paper are: (1) We proposed a novel Crisscross Injection Strategy (CIS) with Boundary Enhance Conditioner (BEC) and Core Enhance Conditioner (CEC) to enhance the diffusion model's capability to learning features in both edge and non-edge areas. (2) We introduced a Generative Pretrain (GP) method for prostate segmentation to reduce the domain gap between the conditional features and the diffusion model features, improving model convergence. (3) We demonstrated our proposed method achieving SOTA performance on three MRI prostate datasets and one ultrasound prostate dataset under four evaluation metrics.

2 Methods

The architecture of CriDiff is shown in Figure 1. In the first stage, we leveraged DPM's generative power to formulate the segmentation task as a generative problem, developing a model that precisely captures the characteristics of prostate images. In the second stage, the CIS injects both boundary and core features into the pre-trained diffusion model in a crisscross way. To effectively learn boundary and core areas features, we employed the proposed BEC and CEC to separately learn the boundary and core features of prostate images, respectively. Finally, a Gaussian noise is guided by the boundary, core and image feature information to generate the final prediction map.

2.1 Generative Pre-train Approach

To reduce the domain gap between the conditional features and the diffusion model features, we introduced a generative pre-train approach for prostate segmentation. The diffusion model operates through two processes: initially, in the forward process, an image is progressively noised over T steps by adding Gaussian noise. Subsequently, in the reverse stage, a neural network learns to recover the original data by reversing this noise addition. Given prostate images I_0, the reverse process can be represented as: $p_\theta(I_{0:T-1} \mid I_T) = \prod_{t=1}^{T} p_\theta(I_{t-1} \mid I_t)$, where θ represents the denoising model parameters and $p_\theta(I_T)$ is the latent variable distribution. Through the training process, the parameters θ acquire the capability to represent prostate features, enabling the transformation from a Gaussian noise distribution to the prostate data distribution $p_\theta(I_0)$. Following [8], the θ is considered as a noise prediction network ϵ_θ, optimized by a simple mean-squared error:

$$\mathcal{L}(\theta) = \|\epsilon_\theta(I_t) - \epsilon_t\|^2, \tag{1}$$

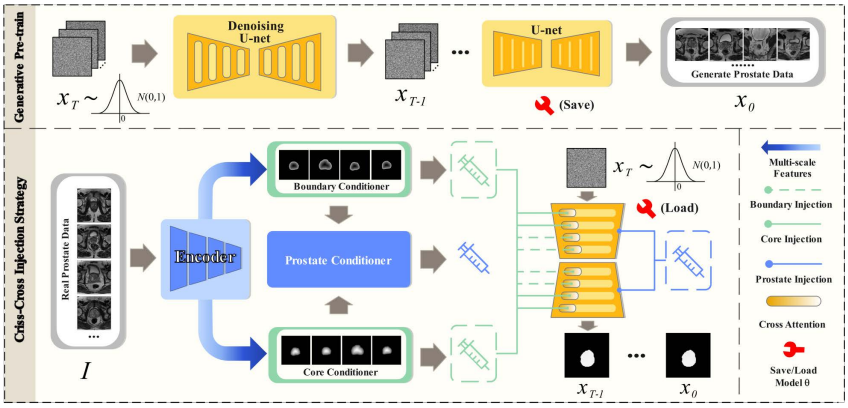

Fig. 1: Overall of our method. Up: The first stage is our proposed generative pre-train that is described in Sec. 2.1. Bottom: After pre-training, we performed the criss-cross injection strategy to segment prostate in Sec. 2.2 and Sec. 2.3.

where I_t is a noised prostate image at t step. By applying the reparameterization, $I_t = \sqrt{\hat{\alpha}_t} I_0 + \sqrt{1-\hat{\alpha}_t}\epsilon_t$, where $\hat{\alpha}$ represents constants hyperparameters, and $\epsilon_t \sim \mathcal{N}(0, \mathbb{I})$ is noise at t step.

2.2 Boundary Enhance Conditioner and Core Enhance Conditioner

Distinct from previous method that enhances edge learning via a weighted loss function, we proposed two parallel conditioners to decouple the learning of edge and core information. As shown in Figure 2, the BEC starts with a higher number of convolutional layers then decreases as the network goes deep. In contrast, the CEC increases the number of convolutional layers as the network deepens. Given that the sideouts of encoder are denoted as f^1, f^2, f^3, f^4 from large to small. Then these features at each level are transformed in parallel into a same number of dimensions (such as 64 in our implementation) via the $Trans$ layers. These layers follow by a combination of 3×3 convolution, batchnorm and relu. Through these layers, we can obtain unified-channel features $B_0^i = Trans(f^i)$ for i from 1 to 4. In this end, the multi-level features of the BEC at the ith row and jth colum B_j^i can be denoted as:

$$B_j^i = \begin{cases} BConv(B_{j-1}^i \copyright Up(B_j^{i+1})), & if\ i+j \leq 4, \\ BConv(B_{j-1}^i \copyright A), if\ i+j=5 \,\&\, i=4, A=0; else\ A = Up(B_{j-1}^{i+1}), \end{cases} \tag{2}$$

where $BConv(\cdot)$ means 3×3 Conv-Bn-Relu operation. © indicates the concatenation operation and $Up(\cdot)$ is an upsample operation with an upsampling rate 2. Analogously, $C_0^i = Trans(f^i)$, $i = 1-4$. The implementation of the CEC at

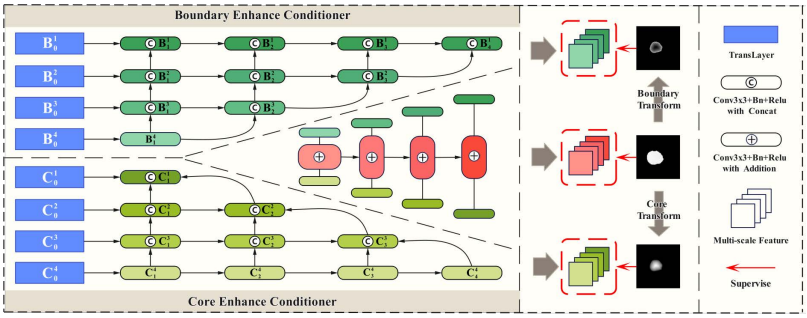

Fig. 2: Detailed structures of our proposed BEC and CEC, which focus on learning the boundary and core information of the prostate under the guidance of decoupled soft labels.

the ith row and jth colum C_j^i formulated as:

$$C_j^i = \begin{cases} BConv(C_{j-1}^i), & if\ i = 4, \\ BConv(C_{j-1}^i \copyright Up(C_j^{i+1}) \copyright A), if\ i < 4 \,\&\, i = j, A = Up(C_{j+1}^{i+1}); \\ & else\ A = 0, \end{cases} \tag{3}$$

then the multi-scale features from the BEC and CEC are fed into a streamlined FPN to obtain the integrated prostate features P^i. It can be defined as:

$$P^i = BConv(B_{5-i}^i \oplus C_i^i \oplus A), if\ i = 4, A = 0; else\ A = Up(P^{i+1}), \tag{4}$$

where $\oplus$ represents a pixel-wise summation operation. Finally, these multi-scale features will be supervised by a combing Dice Loss, BCE Loss, and IoU Loss, thus the total loss of our conditioners is:

$$\mathcal{L}_c = \mathcal{L}_{bce}(B_{5-i}^i, g_b) + \mathcal{L}_{bce}(C_i^i, g_c) + \mathcal{L}_{bce}(P^i, g_p) + \mathcal{L}_{IoU}(P^i, g_p) + \mathcal{L}_{Dice}(P^i, g_p), \tag{5}$$

where g_b and g_c denote the boundary label and the core label of the prostate label g_p. We apply the Distance Transformation (DT) [11] on g_p to differentiate between g_b and g_c, obtaining a gradient image I'. After normalization to [0,1] range, pixels within the object's center exhibit the highest values, while those distant from the center or within the background display the lowest values. Consequently, I' reflects the central, more easily distinguishable aspects of the original image. We then define the core label and the boundary label as $g_c = g_p * I'$, $g_b = g_p * (1 - I')$, respectively.

2.3 Crisscross Injection Strategy

The proposed BEC and CEC are capable of capturing boundary and core features. However, directly injecting these features into the diffusion model in a stage-by-stage manner results in suboptimal feature utilization. Thus, we proposed a crisscross injection strategy that allows the diffusion model to focus on

Table 1: Quantitative comparisons of DSC, IoU, HSD and ASD on three MRI datasets and one ultrasound datasets. For brevity, we denoted these metrics as D, I, H, and A, respectively. The top two results are marked in *red*, blue.

Method/Years	NCI-ISBI [5]				ProstateX [12]				Promise12 [13]				CCH-TRUSPS [6]			
	$D\uparrow$	$I\uparrow$	$H\downarrow$	$A\downarrow$	$D\uparrow$	$I\uparrow$	$H\downarrow$	$A\downarrow$	$D\uparrow$	$I\uparrow$	$H\downarrow$	$A\downarrow$	$D\uparrow$	$I\uparrow$	$H\downarrow$	$A\downarrow$
Unet [16]$_{15MICCAI}$	.822	.786	2.32	3.99	.748	.683	3.41	3.93	.779	.676	4.25	6.57	.898	.848	5.58	7.43
Unet++ [27]$_{19TMI}$	.814	.777	2.30	3.76	.741	.682	3.44	3.80	.810	.715	4.12	5.06	.882	.824	5.92	7.80
TransUnet [4]$_{21Arxiv}$	.827	.789	2.28	3.90	.851	.795	2.92	2.33	.887	.812	3.65	2.22	.915	.874	5.36	4.43
Swin-Unet [3]$_{22ECCV}$	.821	.782	2.39	5.01	.792	.727	3.32	3.49	.839	.744	4.09	3.60	.908	.857	5.78	5.83
Uctransnet [19]$_{22AAAI}$	.813	.776	2.38	4.86	.769	.701	3.40	2.93	.875	.796	3.87	4.62	.915	.868	5.61	5.39
G-CASCADE [15]$_{24WACV}$	.842	.808	2.24	3.75	.844	.795	3.05	2.02	.880	.802	3.67	2.67	.915	.871	5.58	5.93
CAT-Net [9]$_{23TMI}$	.841	.810	2.21	4.04	.796	.743	3.31	2.67	.888	.813	3.76	2.51	.895	.850	5.76	5.01
CCT-Unet [25]$_{23JBHI}$	.836	.803	2.20	4.50	.803	.756	3.08	2.22	.857	.775	3.82	3.51	.902	.852	5.64	6.98
MicroSegNet [10]$_{24CMIG}$	.829	.796	2.25	3.86	.849	.798	2.96	2.45	.890	.817	3.66	2.19	.928	.886	5.49	4.72
SegDiff [1]$_{21Arxiv}$	.807	.776	2.27	4.12	.835	.788	3.07	1.93	-	-	-	-	.854	.788	5.86	7.83
EnDiff [22]$_{22MIDL}$	.814	.781	2.32	3.82	.815	.761	3.24	2.17	-	-	-	-	.875	.829	6.04	5.71
DermoSegDiff [2]$_{23MICCAI}$	.841	.806	2.14	3.79	.853	.804	2.96	2.02	.885	.809	3.69	2.64	.900	.855	5.41	4.59
MedSegDiff-V2 [24]$_{24AAAI}$	.828	.796	2.19	3.71	.822	.773	3.10	2.18	.888	.815	3.67	2.19	.844	.772	6.05	8.55
Ours	.858	.827	2.04	3.13	.874	.824	2.86	1.85	.899	.828	3.63	2.06	.923	.883	5.35	4.17

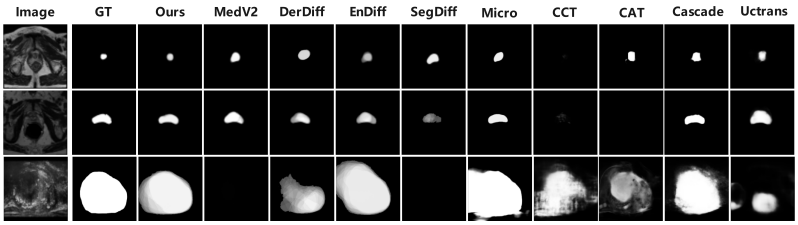

Fig. 3: Visual comparisons of the proposed model and existing SOTA methods.

object localization in early stages and refines the object's edges in later stages. Following [23,24], we adopted a modified ResUNet as our diffusion backbone and inject prostate features into the backbone layer by layer. The encoder contains four convolutional stage with sequentially decreasing resolution. Conversely, the decoder consists of four convolutional stages with sequentially increasing resolution. We applied the cross-attention to facilitate the interaction between both boundary and core features with the diffusion feature. Finally, the proposed Crisscross Injection Strategy can be denoted as:

$$CIS = \begin{cases} E^i = CroAtt(A, E^i),\ if\ i = 1, 2, A = C_i^i;\ else\ A = B_{5-i}^i, \\ D^{5-i} = CroAtt(A, D^{5-i}),\ if\ i = 1, 2, A = C_i^i;\ else\ A = B_{5-i}^i, \end{cases} \tag{6}$$

where E^i means the ith outputs of encoder stage and D^i is the ith outputs of decoder stage, respectively. $CroAtt(\cdot)$ is the cross-attention operataion.

Table 2: Different conditioner settings.

Index	Conditioner				Param Size(M)	ProstateX [12]				CCH-TRUSPS [6]			
	P	P*	C	B		$D\uparrow$	$I\uparrow$	$H\downarrow$	$A\downarrow$	$D\uparrow$	$I\uparrow$	$H\downarrow$	$A\downarrow$
(1)	✓				53.35	.843	.778	3.01	2.46	.895	.856	5.67	5.09
(2)		✓			53.36	.857	.799	2.95	2.21	.906	.867	5.61	4.85
(3)	✓		✓		54.06	.852	.801	2.95	2.26	.909	.871	5.56	4.45
(4)	✓			✓	54.13	.865	.811	2.85	2.09	.916	.873	5.55	4.36
(5)	✓		✓	✓	54.63	.874	.824	2.86	1.85	.923	.883	5.35	4.17

Table 3: Different init method comparison.

Method	ProstateX [12]				CCH-TRUSPS [6]			
	$D\uparrow$	$I\uparrow$	$H\downarrow$	$A\downarrow$	$D\uparrow$	$I\uparrow$	$H\downarrow$	$A\downarrow$
Random	.865	.812	2.92	1.91	.893	.858	5.68	4.72
Kaiming	.868	.817	2.95	1.91	.896	.862	5.63	4.52
Ours	.874	.824	2.86	1.85	.923	.883	5.35	4.17

3 Experiments

3.1 Experiment Protocol

Datasets. We performed the evaluation on four public benchmark datasets, categorized into two types: three datasets comprising MRI images (NCI-ISBI [5], ProstateX [12] and Promise12 [13]) and one dataset consisting of ultrasound images (CCH-TRUSPS [6]). Details of these datasets are provided in the supplementary material. **Metrics.** To validate the proposed model, we adopt four metrics: Dice Similarity Coefficient (DSC), Intersection over Union (IoU), Hausdorff Distance (HSD) and Average Surface Distance (ASD). **Implementation detail.** We trained our network using the Pytorch toolbox on two RTX 4090 GPUs, employing PVT-B2 [20] as the encoder. The training utilized a batch size of 6, the AdamW optimizer with a 1e-5 learning rate, and included 100,000 iterations. We conducted 25 ensemble runs with T=500. Pre-train phase details are in the supplementary material.

3.2 Comparison with State-of-the-arts

The performance of the proposed method is compared with 6 general medical image segmentation methods, including Unet [16], Unet++ [27], TransUnet [4], Swin-Unet [3], Uctransnet [19] and G-CASCADE [15]. We also compared 3 prostate segmentation methods, including CAT-Net [9], CCT-Unet [25] and MicroSegNet [10]. Additionally, we compared 4 DPM-based methods included SegDiff [1], EnDiff [22], DermoSegDiff [2] and MedSegDiff-V2 [24]. For a fair comparison, we replaced the encoder of these DPM-based methods with PVT-B2 [20], except for the EnDiff, which does not utilize the encoder. We trained and inferred these methods for the same number of iterations and ensemble times as our model. **Results.** As shown in Table 1, our method improves the IoU by an average of 2.1% and reduces the ASD by an average of 8.5% across three MRI prostate datasets compared to the second-best method. On the ultrasound prostate dataset, the Dice and IoU metrics show only minor differences at the thousandth digit compared to the best method. This may be due to the simplicity of this dataset, where multiple ensemble runs might introduce noise, slightly reducing performance. Intuitively, We visualize segmentation maps from our model and others in Figure 3. It is obvious that our model not only achieves precise localization but also clearly delineates boundaries for prostates of varying sizes. More visualized results can be found in the supplementary material.

Table 4: Injection strategy comparison.

Index	Strategy				ProstateX [12]				CCH-TRUSPS [6]			
	SbS	2:2	1:3	3:1	$D\uparrow$	$I\uparrow$	$H\downarrow$	$A\downarrow$	$D\uparrow$	$I\uparrow$	$H\downarrow$	$A\downarrow$
(1)	✓				.856	.806	2.95	1.96	.903	.862	5.53	4.73
(2)		✓			.874	.824	2.86	1.85	.923	.883	5.35	4.17
(3)			✓		.875	.820	2.89	1.99	.927	.875	5.45	4.45
(4)				✓	.865	.816	2.91	1.95	.923	.888	5.32	4.26

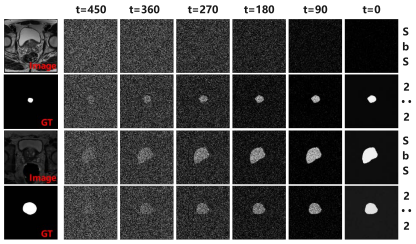

Fig. 4: Visual comparisons with (1) and (2) are shown in Table 4.

3.3 Ablation Study

Effect of BEC and CEC. We validated our proposed conditioner structures through five ablation studies. As shown in Table 2, (1) only prostate conditioner to inject. (2) a simple FPN replaced three conditioners, predicting three-channel features representing prostate, core and boundary features for injecting. (3) prostate and core conditioners to inject. (4) prostate and boundary conditioners to inject. (5) all three conditioners to inject. When injecting the same features, we observed that our proposed boundary and core conditioners significantly enhanced model performance compared with (2) and (5). These results demonstrate decoupled learning of boundary and core features can more effectively improve performance with a smaller model size. Moreover, compared with (1), (3), (4) and (5), the absence of either one or both boundary and core conditioners for learning and injecting features resulted in a decrease in Dice scores on the ProstateX and CCH-TRUSPS datasets, respectively. This further highlights the effectiveness of the boundary and core conditioner architectures.

Effect of CIS. We performed four quantitative experiments to validate the effectiveness of CIS. As shown in Table 4, 2:2 means a ratio of injection layer for using the proposed strategy to inject core features into the shallow two layers and detail features into the deeper two layers of both encoder and decoder. SbS denotes that the stage by stage strategy injects boundary features and core features into shallow layers and deeper layers with a 2:2 ratio. Compared (1) with (2), (3), and (4), we observed that our proposed injection strategy significantly outperforms the traditional stage-by-stage injection approach, regardless of whether using a 3:1 or 1:3 ratio. These results strongly validate the adequacy of utilizing multi-level features in our injection strategy. To further illustrate the qualitative effect of our strategy, we visualized feature maps of (1) and (2) in Figure 4. It can be seen that the proposed strategy enables the diffusion model to focus on object localization, preventing entirely black predictions, especially when the prostate region is small. Simultaneously, it allows the model to focus on edge areas achieving precise edge segmentation.

Effect of GP. To validate the effectiveness of GP, we conducted a set of experiments over different initialization methods (Random and Kaiming). As shown in Table 3, our method demonstrates superior performance across two datasets on four metrics, especially showing 3.3% and 2.8% improvements in Dice and IoU

on the CCH-TRUSPS dataset, thereby affirming the benefits of generative pre-training for model initialization. Through the proposed GP, the diffusion model acquire the capability to represent prostate features, bridging the domain gap between the conditional features and the diffusion model features. To illustrate this point, we showcased some generated prostate images in the supplementary material. These images clearly possess structures characteristic of prostate imagery and closely resemble real prostate images.

4 Conclusion

In this paper, we proposed CriDiff, a novel framework for prostate segmentation, which efficiently learns and injects multi-scale edge and non-edge features into the diffusion network using two parallel conditioners (BEC and CEC) and a crisscross injection strategy (CIS). To bridge the domain gap between image and diffusion model features, we introduced a generative method without introducing additional parameters. Experimental results demonstrate that our proposed method can achieve state-of-the-art performance in prostate segmentation.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (62376050, 62372080, 62172070, and U22B2052), the Dalian Science and Technology Innovation Foundation (2023JJ11CG001 and 2022JJ11CG001), and the CAAI-Huawei MindSpore Open Fund (CAAIXSJLJJ-2022-014C).

Disclosure of Interests. We declared no competing interests.

References

1. Amit, T., Shaharbany, T., Nachmani, E., Wolf, L.: Segdiff: Image segmentation with diffusion probabilistic models. arXiv preprint arXiv:2112.00390 (2021) 1, 3.2
2. Bozorgpour, A., Sadegheih, Y., Kazerouni, A., Azad, R., Merhof, D.: Dermosegdiff: A boundary-aware segmentation diffusion model for skin lesion delineation. In: International Workshop on PRedictive Intelligence In MEdicine. pp. 146–158. Springer (2023) 1, 1, 3.2
3. 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: European conference on computer vision. pp. 205–218. Springer (2022) 1, 3.2
4. Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A.L., Zhou, Y.: Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306 (2021) 1, 3.2
5. Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., et al.: The cancer imaging archive (tcia): maintaining and operating a public information repository. Journal of digital imaging **26**, 1045–1057 (2013) 1, 3.1
6. Feng, Y., Atabansi, C.C., Nie, J., Liu, H., Zhou, H., Zhao, H., Hong, R., Li, F., Zhou, X.: Multi-stage fully convolutional network for precise prostate segmentation in ultrasound images. Biocybernetics and Biomedical Engineering **43**(3), 586–602 (2023) 1, 2, 3, 3.1, 4

7. Guo, Y., Gao, Y., Shen, D.: Deformable mr prostate segmentation via deep feature learning and sparse patch matching. IEEE transactions on medical imaging **35**(4), 1077–1089 (2015) 1
8. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020) 2.1
9. Hung, A.L.Y., Zheng, H., Miao, Q., Raman, S.S., Terzopoulos, D., Sung, K.: Catnet: A cross-slice attention transformer model for prostate zonal segmentation in mri. IEEE transactions on medical imaging **42**(1), 291–303 (2023) 1, 3.2
10. Jiang, H., Imran, M., Muralidharan, P., Patel, A., Pensa, J., Liang, M., Benidir, T., Grajo, J.R., Joseph, J.P., Terry, R., et al.: Microsegnet: A deep learning approach for prostate segmentation on micro-ultrasound images. Computerized Medical Imaging and Graphics p. 102326 (2024) 1, 3.2
11. Kimmel, R., Kiryati, N., Bruckstein, A.M.: Sub-pixel distance maps and weighted distance transforms. Journal of Mathematical Imaging and Vision **6**, 223–233 (1996) 2.2
12. Litjens, G., Debats, O., Barentsz, J., Karssemeijer, N., Huisman, H.: Computer-aided detection of prostate cancer in mri. IEEE transactions on medical imaging **33**(5), 1083–1092 (2014) 1, 2, 3, 3.1, 4
13. Litjens, G., Toth, R., Van De Ven, W., Hoeks, C., Kerkstra, S., Van Ginneken, B., Vincent, G., Guillard, G., Birbeck, N., Zhang, J., et al.: Evaluation of prostate segmentation algorithms for mri: the promise12 challenge. Medical image analysis **18**(2), 359–373 (2014) 1, 3.1
14. Pellicer-Valero, O.J., Marenco Jimenez, J.L., Gonzalez-Perez, V., Casanova Ramon-Borja, J.L., Martin Garcia, I., Barrios Benito, M., Pelechano Gomez, P., Rubio-Briones, J., Rupérez, M.J., Martín-Guerrero, J.D.: Deep learning for fully automatic detection, segmentation, and gleason grade estimation of prostate cancer in multiparametric magnetic resonance images. Scientific reports **12**(1), 2975 (2022) 1
15. Rahman, M.M., Marculescu, R.: G-cascade: Efficient cascaded graph convolutional decoding for 2d medical image segmentation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 7728–7737 (2024) 1, 3.2
16. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. pp. 234–241. Springer (2015) 1, 3.2
17. Siegel, R.: Cancer statistics, 2020. CA: a cancer journal for clinicians. **70**(1), 7 (2020) 1
18. Tian, Z., Liu, L., Zhang, Z., Fei, B.: Superpixel-based segmentation for 3d prostate mr images. IEEE transactions on medical imaging **35**(3), 791–801 (2015) 1
19. Wang, H., Cao, P., Wang, J., Zaiane, O.R.: Uctransnet: rethinking the skip connections in u-net from a channel-wise perspective with transformer. In: Proceedings of the AAAI conference on artificial intelligence. vol. 36, pp. 2441–2449 (2022) 1, 3.2
20. Wang, W., Xie, E., Li, X., Fan, D.P., Song, K., Liang, D., Lu, T., Luo, P., Shao, L.: Pvtv2: Improved baselines with pyramid vision transformer. Computational Visual Media **8**(3), 1–10 (2022) 3.1, 3.2
21. Warfield, S.K., Zou, K.H., Wells, W.M.: Validation of image segmentation and expert quality with an expectation-maximization algorithm. In: Medical Image Computing and Computer-Assisted InterventionMICCAI 2002: 5th International

Conference Tokyo, Japan, September 25–28, 2002 Proceedings, Part I 5. pp. 298–306. Springer (2002) 1
22. Wolleb, J., Sandkühler, R., Bieder, F., Valmaggia, P., Cattin, P.C.: Diffusion models for implicit image segmentation ensembles. In: International Conference on Medical Imaging with Deep Learning. pp. 1336–1348. PMLR (2022) 1, 1, 3.2
23. Wu, J., Fu, R., Fang, H., Zhang, Y., Xu, Y.: Medsegdiff-v2: Diffusion based medical image segmentation with transformer. arXiv preprint arXiv:2301.11798 (2023) 1, 2.3
24. Wu, J., FU, R., Fang, H., Zhang, Y., Yang, Y., Xiong, H., Liu, H., Xu, Y.: Medsegdiff: Medical image segmentation with diffusion probabilistic model. In: Medical Imaging with Deep Learning (2023) 1, 1, 2.3, 3.2
25. Yan, Y., Liu, R., Chen, H., Zhang, L., Zhang, Q.: Cct-unet: A u-shaped network based on convolution coupled transformer for segmentation of peripheral and transition zones in prostate mri. IEEE Journal of Biomedical and Health Informatics (2023) 1, 3.2
26. Yu, L., Yang, X., Chen, H., Qin, J., Heng, P.A.: Volumetric convnets with mixed residual connections for automated prostate segmentation from 3d mr images. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 31 (2017) 1
27. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J.: Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. IEEE transactions on medical imaging **39**(6), 1856–1867 (2019) 1, 3.2