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

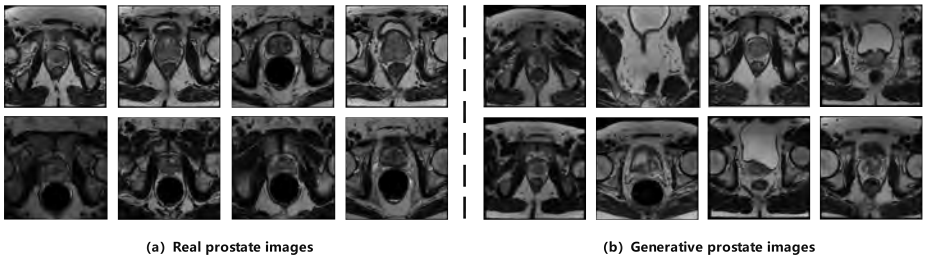

Fig. 1: Some generative prostate images compare with real prostate images.

1 Detailed of Experiments

1.1 Datasets

We performed the evaluation on four public benchmark datasets. NCI-ISBI is from ISBI 2013 Prostate Magnetic Resonance Imaging Challenge, containing 1571 T2 weighted images (T2WI) for training, 271 T2WI for testing. ProstateX includes 664 T2WI for training and 166 T2WI for testing, which are performed by Siemens MAGNETOM Trio and Skyra 3T MR scanners. Promise12 is from the Prostate MR Image Segmentation challenge, containing 778 T2WI for training and 418 T2WI for testing. CCH-TRUSPS is an ultrasound prostate segmentation dataset collected from Chongqing University Cancer Hospital, containing 2152 ultrasound train images and 727 ultrasound test images. All 3D scans are converted into 2D slices. Then, each slice is resized to 256×256 and normalized to $[0, 1]$ for training.

1.2 Detail of Pre-train Stage

Following the settings of most DPM-based methods, we adopted a ResUNet without linear attention layers as the backbone for diffusion and used DDPM sampling strategy to generate images both two stages. We used the Adam optimization with an initial learning rate of 8e-5 to optimize the architecture parameters θ. For each dataset, we utilized the training sets from each of the four

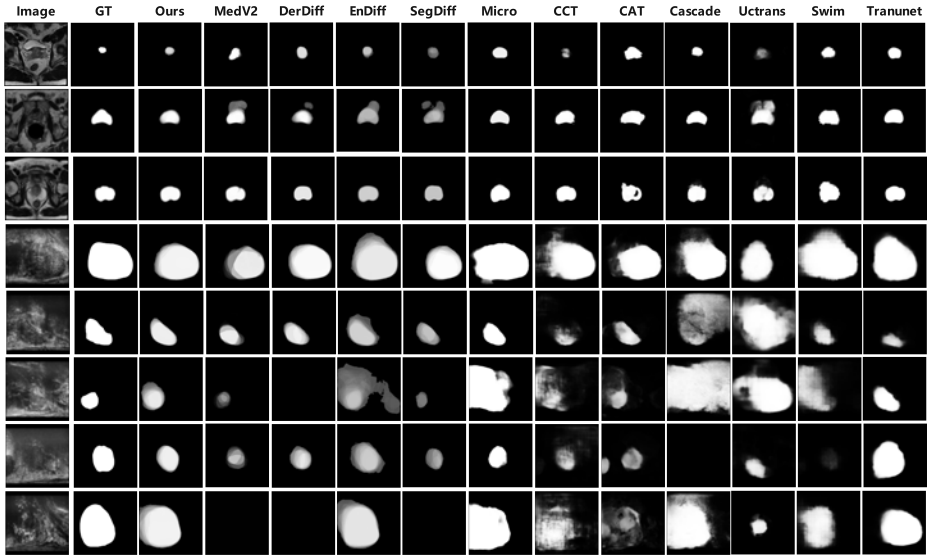

Fig. 2: More visual comparisons of the proposed model and existing SOTA methods.

Table 1: Different methods inject feature into the Diffusion.

Index	Concat	CroAtt	Param Size(M)	ProstateX				CCH-TRUSPS			
				$D\uparrow$	$I\uparrow$	$H\downarrow$	$A\downarrow$	$D\uparrow$	$I\uparrow$	$H\downarrow$	$A\downarrow$
(1)	✓		58.86	.861	.801	2.99	1.84	.913	.875	5.46	4.41
(2)		✓	54.63	.874	.824	2.86	1.85	.923	.883	5.35	4.17

datasets mentioned above as inputs, training the network with 24 batch-size and 20000 iterations with data augmentation (random flipping and cropping). Finally, the pre-trained network θ will be load in the second stage.

2 More Results

2.1 Generative Prostate images

We displayed some images comparisons between prostate images generated using ProstateX dataset and the real images from the same dataset as shown in Figure 1. It is evident that the prostate images created with DPM-based techniques possess a comprehensive structure, and some images appearing remarkably lifelike.

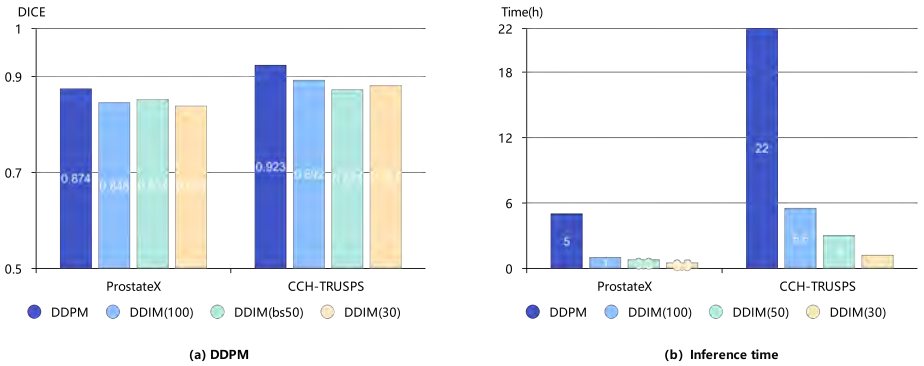

Fig. 3: Comparisons with DDPM and DDIM sampling strategy in Dice and time metrics on two datasets. We set 5 batchsize in DDPM and DDIM with different number of step (100, 50 and 30 respectively).

2.2 More Visual Comparisons

We provided additional visual comparisons between our method and state-of-the-art (SOTA) approaches as shown in 2. These results demonstrate that our method performs well not only on MRI, achieving precise edge segmentation, but also offers improved localization in ultrasound images. This further proves the effectiveness of our approach.

2.3 Crisscross Injection Strategy

Existing DPM-based models inject features into the diffusion backbone mainly with 2 operations by concatenation and cross-attention. We conducted 2 experiments that utilize the proposed injection strategy with different operations to validate the performance of these two operations as shown in Table 1. These results are shown that the concatenation operation not only increases the model's parameter but also reduces its performance. Thus, we adopted the cross-attention operation to inject boundary features and core features.

3 Limitation

Despite the good results of our approach, there is still a limitation. Currently, most DPM-based methods for segmentation mainly adopt DDPM and DDIM sampling strategies. DDIM can adopt the same training optimization objective as DDPM, but it can adopt fewer steps in sampling to improve the speed. We performed a set of experiments to validate the performance of difference sampling strategy as shown in Figure 3. These results indicate the DDIM sampling strategy is fast, but the accuracy cannot be guaranteed, and it needs laborious trial and error to determine the number of sampling steps. While the DDPM has a high

accuracy, but the inference is slow, which seriously hinders its application on other tasks. Exploring a fast and accurate sampling strategy that we aim to address in future work.