Diffusion Adversarial Representation Learning for Self-supervised Vessel Segmentation

ICLR 2023

https://github.com/bispl-kaist/DARL

李响 | 2023.3.25

目录

CONTENTS

01. Problems and contributions

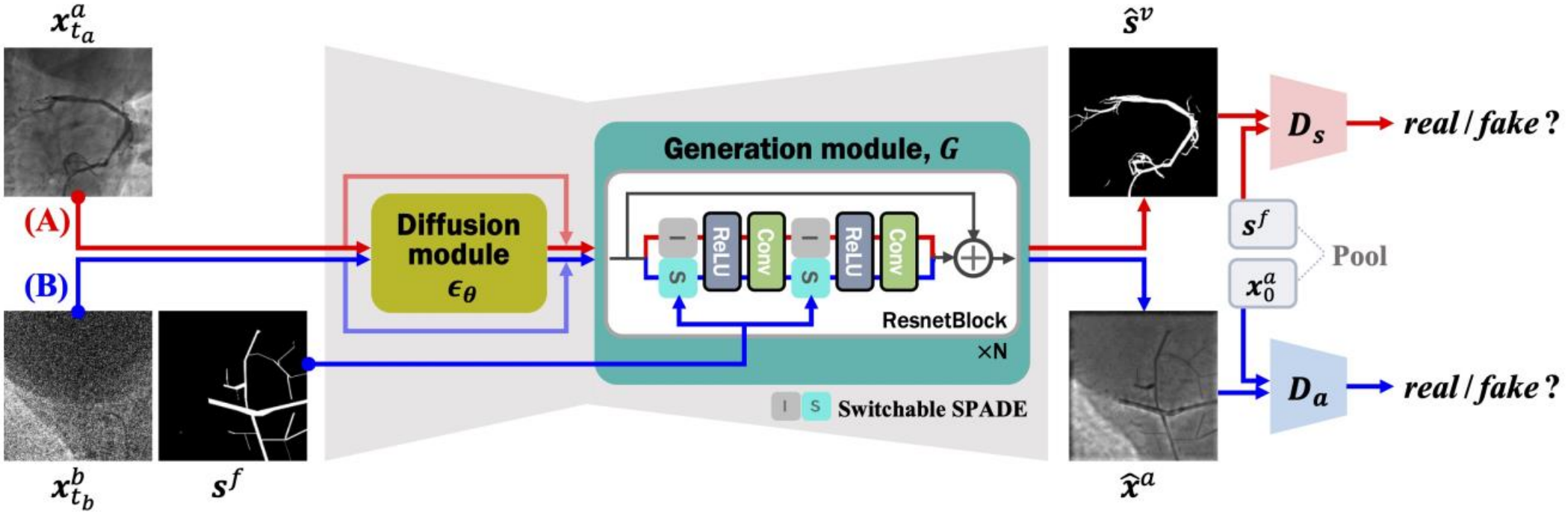
Problems:

- 数据量极小。
- 血管图像背景复杂、对比度低、有运动伪影和许多微小的分支。

Contributions:

- 提出了一种非迭代式扩散模型，用于图像生成和自监督血管分割。
- 提出了一种具有可切换 SPADE（spatially adaptive denormalization）层的生成模块，该模块能够生成合成的血管造影图像，并分割血管结构。
- 扩散对抗表示学习（Diffusion Adversarial Representation Learning, DARL）模型，该模型利用降噪扩散概率模型（DDPM）和对抗学习进行血管分割。不需要大量的真实标签，并且可以承受严重的噪声，适用于多模态。在各种数据集上优于现有的无监督和自监督血管分割方法。

02. Overview



Self-supervised vessel segmentation:

由于血管结构复杂，很难获得细致的标签用于监督学习。虽然使用无监督学习可以缓解标签稀缺问题，但是用于分割细小的血管并获得合理性能的完全无监督的方法相对较少。作者提出了一种自监督学习的方法，该方法可以利用数据本身生成的自监督标签有效地学习目标表示信息。

03. Denoising diffusion probabilistic model

$$q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) = \mathcal{N}(\boldsymbol{x}_t; \sqrt{1-\beta_t}\boldsymbol{x}_{t-1}, \beta_t\boldsymbol{I}),$$

其中，βt 是指固定方差 (0~1)。在采样得到 xt 的时候并不是直接通过高斯分布 q(xt|xt-1) 采样，而是用了一个重参数化的技巧。

前向扩散过程还有个属性，可以直接从 x0 采样得到中间任意一个时间步的噪声图片 xt，所以可以写成：

$$q(\boldsymbol{x}_t|\boldsymbol{x}_0) = \mathcal{N}(\boldsymbol{x}_t; \sqrt{\alpha_t}\boldsymbol{x}_0, (1-\alpha_t)\boldsymbol{I}),$$

where $\alpha_t = \Pi_{s=1}^{t}(1-\beta_s)$. Then, DDPM is trained to approximate reverse diffusion process:

$$p_\theta(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t) = \mathcal{N}(\boldsymbol{x}_{t-1}; \boldsymbol{\mu}_\theta(\boldsymbol{x}_t, t), \sigma_t^2\boldsymbol{I}),$$

where σ_t is a fixed variance, and $\boldsymbol{\mu}_\theta$ is a parameterized mean with the noise predictor ϵ_θ:

$$\boldsymbol{\mu}_\theta(\boldsymbol{x}_t, t) = \frac{1}{\sqrt{1-\beta_t}}\left(\boldsymbol{x}_t - \frac{\beta_t}{\sqrt{1-\alpha_t}}\boldsymbol{\epsilon}_\theta(\boldsymbol{x}_t, t)\right).$$

Thus, in the generative process, the sample can be obtained from the Gaussian noise by the iterative denoising steps: $\boldsymbol{x}_{t-1} = \boldsymbol{\mu}_\theta(\boldsymbol{x}_t, t) + \sigma_t\boldsymbol{z}$, where $\boldsymbol{z} \sim \mathcal{N}(0, I)$.

03. Generation module with switchable SPADE layer

$$\boldsymbol{v} = \begin{cases} \mathrm{SPADE}(\boldsymbol{v}, \boldsymbol{s}), & \text{if mask } \boldsymbol{s} \text{ is given,} \\ \mathrm{IN}(\boldsymbol{v}), & \text{otherwise,} \end{cases}$$

of

where IN is the instance normalization (Ulyanov et al., 2017). So, when model is given the fake vessel mask s^f, the SPADE is computed by:

$$x_{m,c,h,w} = \gamma_{c,h,w}(\boldsymbol{s}^f)\frac{x_{m,c,h,w} - \mu_c}{\sigma_c} + \beta_{c,h,w}(\boldsymbol{s}^f),$$

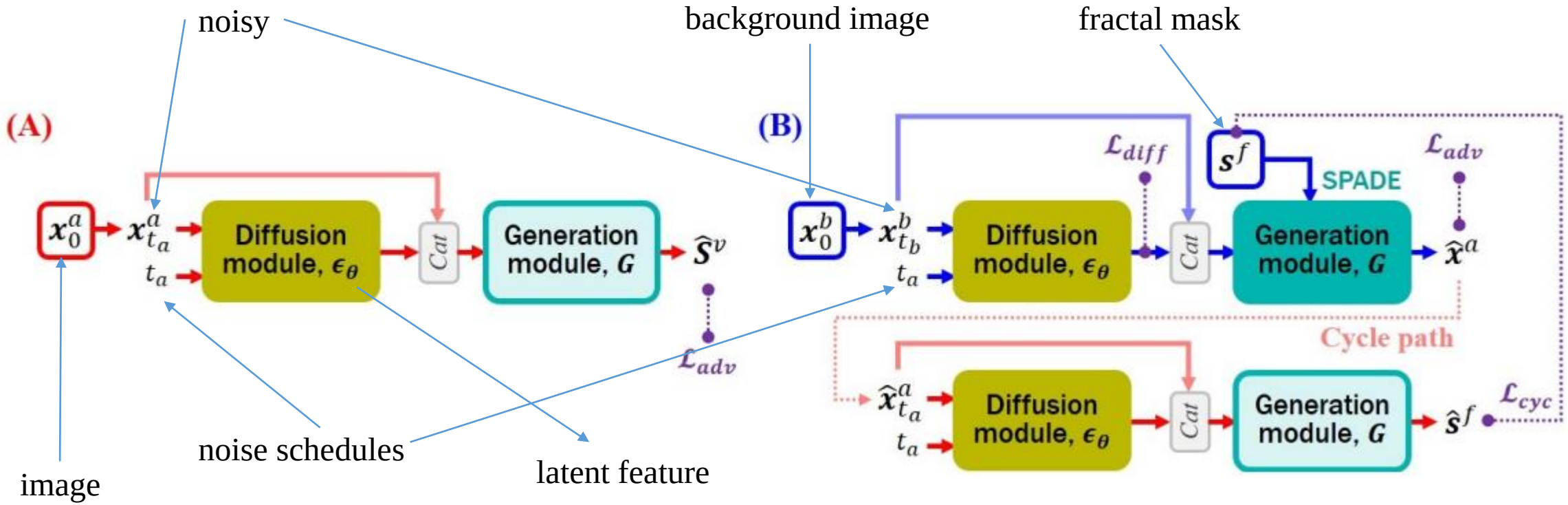
显示了如何使用背景信号的平均值 (μ) 和标准差 (σ) 对由 x 表示的输入图像数据进行归一化。对图像的每个通道 (c) 和空间位置 (h, w) 分别进行归一化，并由两个可学习的参数控制：γ 和 β，这两个参数具有空间自适应性，可根据对抗学习过程进行调整。

(A) path: $\hat{\boldsymbol{s}}^v = G(\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}_t^a, t); \mathbf{0})$.　　(B) path: $\hat{\boldsymbol{x}}^a = G(\boldsymbol{\epsilon}_{\boldsymbol{\theta}}(\boldsymbol{x}_t^b, t); \boldsymbol{s}^f)$.

其中，$\boldsymbol{\epsilon}_{\boldsymbol{\theta}}$ 是用 diffusion module 学习的。

归一化有助于减少混乱的背景结构对血管分割的影响，而空间自适应和可切换的参数使模型能够适应不同的图像区域和血管图像大小。

04. NETWORK TRAINING



LOSS FUNCTION:

$$\min_{\boldsymbol{\theta},G} \mathcal{L}^G(\boldsymbol{\epsilon_\theta}, G, D_s, D_a), \quad \min_{D_s,D_a} \mathcal{L}^D(\boldsymbol{\epsilon_\theta}, G, D_s, D_a),$$

where $\mathcal{L}^G$, and $\mathcal{L}^D$ denotes the losses for the diffusion/generator and discriminator, respectively, which are given by:

$$\mathcal{L}^G(\boldsymbol{\epsilon_\theta}, G, D_s, D_a) = \mathcal{L}_{diff}(\boldsymbol{\epsilon_\theta}) + \alpha\mathcal{L}^G_{adv}(\boldsymbol{\epsilon_\theta}, G, D_s, D_a) + \beta\mathcal{L}_{cyc}(\boldsymbol{\epsilon_\theta}, G),$$

$$\mathcal{L}^D(\boldsymbol{\epsilon_\theta}, G, D_s, D_a) = \mathcal{L}^{D_s}_{adv}(\boldsymbol{\epsilon_\theta}, G, D_s) + \mathcal{L}^{D_a}_{adv}(\boldsymbol{\epsilon_\theta}, G, D_a),$$

04. NETWORK TRAINING

Diffusion loss:

$$\mathcal{L}_{diff}(\boldsymbol{\epsilon_\theta}) := \mathbb{E}_{t,\boldsymbol{x}_0,\boldsymbol{\epsilon}}\left[\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\sqrt{\alpha_t}\boldsymbol{x}_0 + \sqrt{1-\alpha_t}\boldsymbol{\epsilon}, t)\|^2\right].$$

Adversarial loss:

$$\mathcal{L}_{adv}^{G}(\boldsymbol{\epsilon_\theta}, G, D_s, D_a) = \mathbb{E}_{\boldsymbol{x}^a}[(D_s(G(\boldsymbol{\epsilon_\theta}(\boldsymbol{x}^a); \mathbf{0})) - 1)^2] + \mathbb{E}_{\boldsymbol{x}^a, \boldsymbol{s}^f}[(D_a(G(\boldsymbol{\epsilon_\theta}(\boldsymbol{x}^b); \boldsymbol{s}^f)) - 1)^2].$$

$$\mathcal{L}_{adv}^{D_s}(\boldsymbol{\epsilon_\theta}, G, D_s) = \frac{1}{2}\mathbb{E}_{\boldsymbol{s}^f}[(D_s(\boldsymbol{s}^f) - 1)^2] + \frac{1}{2}\mathbb{E}_{\boldsymbol{x}^a}[(D_s(G(\boldsymbol{\epsilon_\theta}(\boldsymbol{x}^a); \mathbf{0})))^2],$$

$$\mathcal{L}_{adv}^{D_a}(\boldsymbol{\epsilon_\theta}, G, D_a) = \frac{1}{2}\mathbb{E}_{\boldsymbol{x}_0^a}[(D_a(\boldsymbol{x}_0^a) - 1)^2] + \frac{1}{2}\mathbb{E}_{\boldsymbol{x}^a, \boldsymbol{s}^f}[(D_a(G(\boldsymbol{\epsilon_\theta}(\boldsymbol{x}^b); \boldsymbol{s}^f)))^2].$$

Cyclic reconstruction loss:

$$\mathcal{L}_{cyc}(\boldsymbol{\epsilon_\theta}, G) = \mathbb{E}_{\boldsymbol{x}^b, \boldsymbol{s}^f}[||G(\boldsymbol{\epsilon_\theta}(G(\boldsymbol{\epsilon_\theta}(\boldsymbol{x}^b); \boldsymbol{s}^f)); \mathbf{0}) - \boldsymbol{s}^f||_1].$$

Interface:

模型在推理过程中不需要迭代反向过程，这与传统的扩散模型不同。意味着，一旦训练了 DARL 模型，就可以从 (A) 路径中单步获得血管分割掩码。

05. DETAILS OF DATASET and NETWORK ARCHITECTURE

Table 6: Detailed dataset for training each (A) segmentation and (B) generation path.

	Input	Train	Validation	Test				
		XCAD	XCAD	XCAD	134 XCA	30 XCA	DRIVE	STARE
(A) Segmentation	Angiogram; $\boldsymbol{x}^a$	1,621	12	114	134	30	20	20
	Ground-truth mask	0	12	114	134	30	20	20
(B) Generation	Background; $\boldsymbol{x}^b$	1,621	-	-	-	-	-	-
	Fractal mask; $\boldsymbol{s}^f$	1,621	-	-	-	-	-	-

Ma, Yuxin, et al. "Self-supervised vessel segmentation via adversarial learning." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021.

Table 12: Training complexity (FLOPS).

Method	(A) path	(B) path	Cycle path	Discriminator	Total
DA	121.80	121.80	121.80 × 2	6.24 × 2	499.68
SSVS	121.80	121.80	121.80 × 2	6.24 × 2	499.68
Ours	90.66	173.51	90.66 × 1	6.24 × 2	367.31

Table 4: Detailed network architecture of the diffusion module. For each block (blk), $C_{i,j}$ is the convolution layer with $i \times i$ kernel and stride length of j, RS_i pairs are entry points for residual shortcut path within a block unit, RB is the residual block module, and SA is the self-attention module. GN is the group normalization, and Ch indicates the size of output channel dimension.

Blk	**Diffusion module**													**Ch**
	Downstream							Upstream						
1	$C_{3,1}$	RS_1	RB	RS_2	RB	RS_3		RS_3	RB	RS_2	RB	RS_1	RB	CB	1
2	$C_{3,2}$	RS_1	RB	RS_2	RB	RS_3		RS_3	RB	RS_2	RB	RS_1	RB	UP	64
3	$C_{3,2}$	RS_1	RB	RS_2	RB	RS_3		RS_3	RB	RS_2	RB	RS_1	RB	UP	128
4	$C_{3,2}$	RS_1	RB	RS_2	RB	RS_3		RS_3	RB	RS_2	RB	RS_1	RB	UP	128
5	$C_{3,2}$	RS_1	RB SA	RS_2	RB SA	RS_3		RS_3	RB SA	RS_2	RB SA	RS_1	RB SA	UP	256
6	$C_{3,2}$	RS_1	RB	RS_2	RB	RS_3		RS_3	RB	RS_2	RB	RS_1	RB	UP	256
Mid					RB		SA		RB						

Note: RB = [RS_n - GN - Swish - $C_{3,1}$ - GN - Swish - $C_{3,1}$ - RS_n], SA = [GN - C_1 - C_1], UP = [Upsample - $C_{3,1}$], CB = [GN - Swich - C_3]

Table 5: Detailed network architecture of the generation module. UP is the nearest neighbor upsampling function, RS_i pairs are entry points for residual shortcut path, $C_{i,j}$ is the convolution layer with $i \times i$ kernel and stride length of j, and IN is the instance normalization layer. S-SPADE is the proposed switchable SPADE layer that turns on SPADE if the semantic layout is provided, otherwise turns off SPADE and applies IN. Ch indicates the size of output channel dimension.

Stream	**Generation module**									**Ch**
	Scale		Conv.	Act.	Norm.	Conv.	Act.	Norm.		
In			C_7		IN		ReLU			64
DownBlock1			$C_{3,2}$		IN		ReLU			128
DownBlock2			$C_{3,2}$		IN		ReLU			256
MidResBlock1		RS_1	$C_{3,1}$	ReLU	S-SPADE	$C_{3,1}$	ReLU	S-SPADE	RS_1	256
MidResBlock2		RS_2	$C_{3,1}$	ReLU	S-SPADE	$C_{3,1}$	ReLU	S-SPADE	RS_2	256
UpResBlock1	UP	RS_3	$C_{3,1}$	ReLU	S-SPADE	$C_{3,1}$	ReLU	S-SPADE	RS_3	128
UpResBlock2	UP	RS_4	$C_{3,1}$	ReLU	S-SPADE	$C_{3,1}$	ReLU	S-SPADE	RS_4	64
Out			C_7							1

05. Result

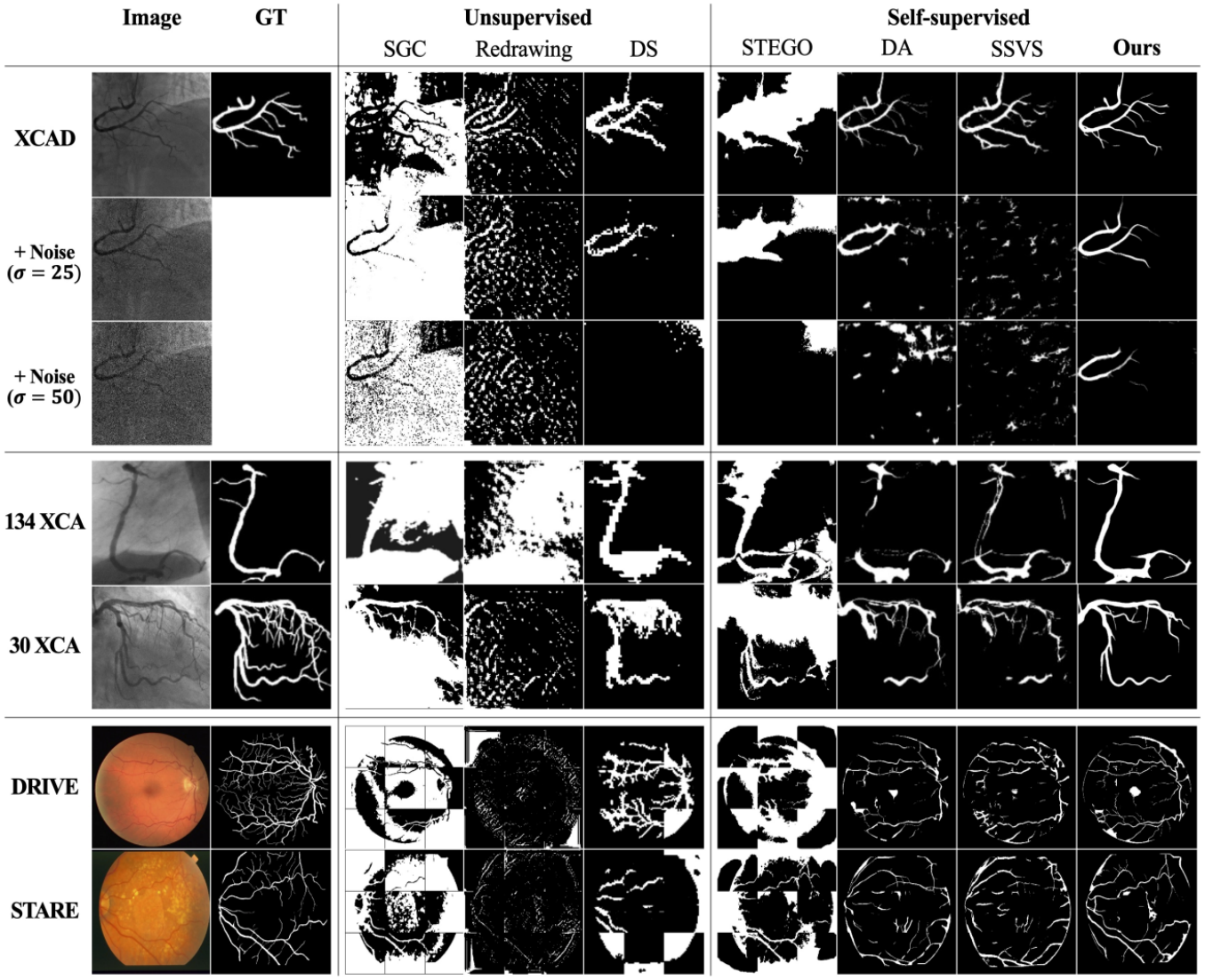


Figure 4: Visual comparison results on the vessel segmentation of various angiography images.

Table 1: Quantitative evaluation results on the vessel segmentation of various angiography images.

Data	Metric	Unsupervised			Self-supervised			
		SGC	Redrawing	DS	STEGO	DA	SSVS	**Ours**
XCAD	IoU	$0.060_{\pm 0.034}$	$0.059_{\pm 0.032}$	$0.366_{\pm 0.105}$	$0.146_{\pm 0.070}$	$0.375_{\pm 0.066}$	$0.410_{\pm 0.087}$	$\mathbf{0.471_{\pm 0.076}}$
	Dice	$0.111_{\pm 0.060}$	$0.109_{\pm 0.056}$	$0.526_{\pm 0.131}$	$0.249_{\pm 0.103}$	$0.542_{\pm 0.073}$	$0.575_{\pm 0.091}$	$\mathbf{0.636_{\pm 0.072}}$
	Precision	$0.062_{\pm 0.034}$	$0.139_{\pm 0.081}$	$0.469_{\pm 0.127}$	$0.152_{\pm 0.077}$	$0.557_{\pm 0.115}$	$0.590_{\pm 0.119}$	$\mathbf{0.701_{\pm 0.115}}$
External test: Coronary angiography								
134 XCA	IoU	$0.045_{\pm 0.035}$	$0.056_{\pm 0.018}$	$0.256_{\pm 0.110}$	$0.134_{\pm 0.081}$	$0.190_{\pm 0.155}$	$0.318_{\pm 0.128}$	$\mathbf{0.426_{\pm 0.059}}$
	Dice	$0.085_{\pm 0.063}$	$0.105_{\pm 0.033}$	$0.394_{\pm 0.159}$	$0.228_{\pm 0.109}$	$0.291_{\pm 0.217}$	$0.468_{\pm 0.156}$	$\mathbf{0.595_{\pm 0.058}}$
	Precision	$0.047_{\pm 0.036}$	$0.058_{\pm 0.019}$	$0.280_{\pm 0.123}$	$0.136_{\pm 0.088}$	$0.506_{\pm 0.201}$	$0.592_{\pm 0.125}$	$\mathbf{0.781_{\pm 0.118}}$
30 XCA	IoU	$0.083_{\pm 0.039}$	$0.048_{\pm 0.022}$	$0.339_{\pm 0.086}$	$0.191_{\pm 0.072}$	$0.298_{\pm 0.109}$	$0.324_{\pm 0.146}$	$\mathbf{0.427_{\pm 0.184}}$
	Dice	$0.150_{\pm 0.064}$	$0.091_{\pm 0.040}$	$0.499_{\pm 0.113}$	$0.314_{\pm 0.100}$	$0.447_{\pm 0.148}$	$0.468_{\pm 0.193}$	$\mathbf{0.572_{\pm 0.205}}$
	Precision	$0.090_{\pm 0.041}$	$0.144_{\pm 0.074}$	$0.525_{\pm 0.130}$	$0.200_{\pm 0.081}$	$0.612_{\pm 0.174}$	$0.613_{\pm 0.212}$	$\mathbf{0.729_{\pm 0.152}}$
Cross-modality test: Retinal imaging								
DRIVE	IoU	$0.063_{\pm 0.055}$	$0.057_{\pm 0.033}$	$0.217_{\pm 0.143}$	$0.152_{\pm 0.073}$	$0.245_{\pm 0.090}$	$0.314_{\pm 0.101}$	$\mathbf{0.372_{\pm 0.148}}$
	Dice	$0.115_{\pm 0.093}$	$0.105_{\pm 0.059}$	$0.333_{\pm 0.201}$	$0.257_{\pm 0.106}$	$0.386_{\pm 0.117}$	$0.469_{\pm 0.119}$	$\mathbf{0.525_{\pm 0.161}}$
	Precision	$0.069_{\pm 0.061}$	$0.199_{\pm 0.155}$	$0.243_{\pm 0.175}$	$0.169_{\pm 0.100}$	$0.503_{\pm 0.218}$	$0.549_{\pm 0.216}$	$\mathbf{0.617_{\pm 0.271}}$
STARE	IoU	$0.055_{\pm 0.045}$	$0.074_{\pm 0.048}$	$0.180_{\pm 0.141}$	$0.125_{\pm 0.076}$	$0.237_{\pm 0.122}$	$0.311_{\pm 0.148}$	$\mathbf{0.368_{\pm 0.191}}$
	Dice	$0.101_{\pm 0.077}$	$0.134_{\pm 0.080}$	$0.281_{\pm 0.201}$	$0.216_{\pm 0.109}$	$0.367_{\pm 0.167}$	$0.454_{\pm 0.185}$	$\mathbf{0.508_{\pm 0.216}}$
	Precision	$0.058_{\pm 0.047}$	$0.227_{\pm 0.157}$	$0.205_{\pm 0.172}$	$0.135_{\pm 0.092}$	$0.427_{\pm 0.233}$	$0.490_{\pm 0.230}$	$\mathbf{0.537_{\pm 0.280}}$

Table 2: Results of noise robustness test according to the Gaussian noise with σ.

σ	Metric	Unsupervised			Self-supervised			
		SGC	Redrawing	DS	STEGO	DA	SSVS	**Ours**
10	IoU	$0.066_{\pm 0.033}$	$0.052_{\pm 0.031}$	$0.331_{\pm 0.104}$	$0.144_{\pm 0.073}$	$0.353_{\pm 0.065}$	$0.258_{\pm 0.079}$	$\mathbf{0.451_{\pm 0.080}}$
	Dice	$0.122_{\pm 0.059}$	$0.096_{\pm 0.053}$	$0.487_{\pm 0.133}$	$0.245_{\pm 0.107}$	$0.519_{\pm 0.073}$	$0.404_{\pm 0.099}$	$\mathbf{0.617_{\pm 0.076}}$
	Precision	$0.069_{\pm 0.035}$	$0.126_{\pm 0.077}$	$0.480_{\pm 0.135}$	$0.157_{\pm 0.091}$	$0.481_{\pm 0.104}$	$0.477_{\pm 0.117}$	$\mathbf{0.710_{\pm 0.115}}$
25	IoU	$0.069_{\pm 0.035}$	$0.036_{\pm 0.021}$	$0.232_{\pm 0.094}$	$0.118_{\pm 0.064}$	$0.247_{\pm 0.072}$	$0.059_{\pm 0.033}$	$\mathbf{0.389_{\pm 0.088}}$
	Dice	$0.128_{\pm 0.061}$	$0.069_{\pm 0.039}$	$0.366_{\pm 0.132}$	$0.206_{\pm 0.095}$	$0.391_{\pm 0.092}$	$0.109_{\pm 0.058}$	$\mathbf{0.554_{\pm 0.092}}$
	Precision	$0.072_{\pm 0.036}$	$0.095_{\pm 0.058}$	$0.446_{\pm 0.159}$	$0.144_{\pm 0.115}$	$0.371_{\pm 0.106}$	$0.149_{\pm 0.082}$	$\mathbf{0.727_{\pm 0.119}}$
50	IoU	$0.070_{\pm 0.025}$	$0.020_{\pm 0.012}$	$0.077_{\pm 0.065}$	$0.060_{\pm 0.050}$	$0.102_{\pm 0.056}$	$0.021_{\pm 0.013}$	$\mathbf{0.269_{\pm 0.081}}$
	Dice	$0.130_{\pm 0.045}$	$0.040_{\pm 0.022}$	$0.136_{\pm 0.109}$	$0.108_{\pm 0.088}$	$0.180_{\pm 0.091}$	$0.041_{\pm 0.025}$	$\mathbf{0.417_{\pm 0.101}}$
	Precision	$0.072_{\pm 0.026}$	$0.061_{\pm 0.038}$	$0.221_{\pm 0.168}$	$0.076_{\pm 0.067}$	$0.169_{\pm 0.094}$	$0.060_{\pm 0.038}$	$\mathbf{0.716_{\pm 0.147}}$

该模型之所以能够在噪声条件下表现良好，是因为它是通过扰乱输入图像的 diffusion module 训练的，因此即使从噪声数据中也可以非常坚固地分割船舶结构。

05. Result

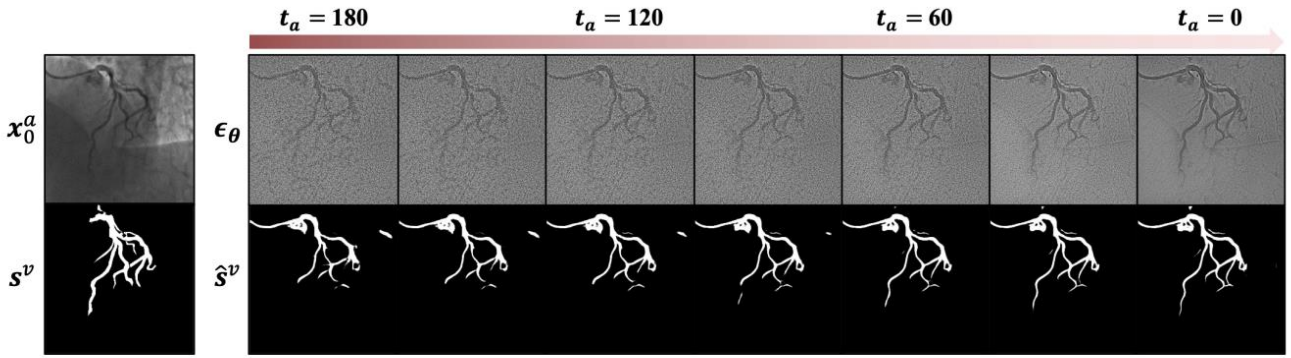


Figure 3: Vessel segmentation according to the noise level t_a. Our model estimates the segmentation masks $\hat{s}^v$ using the latent features ϵ_θ for the noisy angiograms $x_{t_a}^a$. s^v is the ground-truth label.

Table 3: Results of ablation study on the proposed model and loss function.

Method	Module		Loss function			Metric		
	Diffusion	Generation	$\mathcal{L}_{diff}$	$\mathcal{L}_{adv}$	$\mathcal{L}_{cyc}$	IoU	Dice	Precision
Ours	✓	✓	✓	✓	✓	$\mathbf{0.471}_{\pm 0.076}$	$\mathbf{0.636}_{\pm 0.072}$	$\mathbf{0.701}_{\pm 0.115}$
(a)		✓		✓	✓	$0.449_{\pm 0.077}$	$0.616_{\pm 0.074}$	$0.646_{\pm 0.106}$
(b)		w/o S-SPADE		✓	✓	$0.439_{\pm 0.080}$	$0.606_{\pm 0.080}$	$0.620_{\pm 0.111}$
(c)	✓	✓	✓	✓		$0.322_{\pm 0.055}$	$0.485_{\pm 0.064}$	$0.580_{\pm 0.112}$
(d)	✓	✓	✓	✓	L1→CE	$0.346_{\pm 0.084}$	$0.508_{\pm 0.094}$	$0.672_{\pm 0.147}$

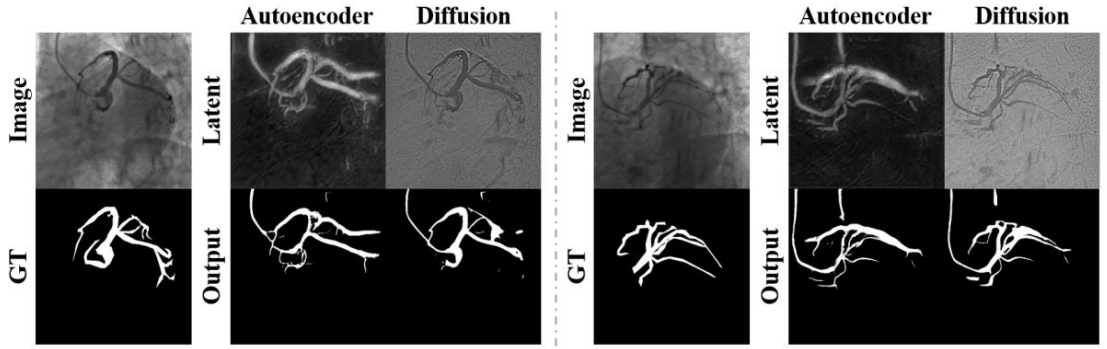


Figure 9: Visual results of vessel segmentation according to the latent feature estimation models.

Table 9: Quantitative evaluation results according to the latent feature estimation models.

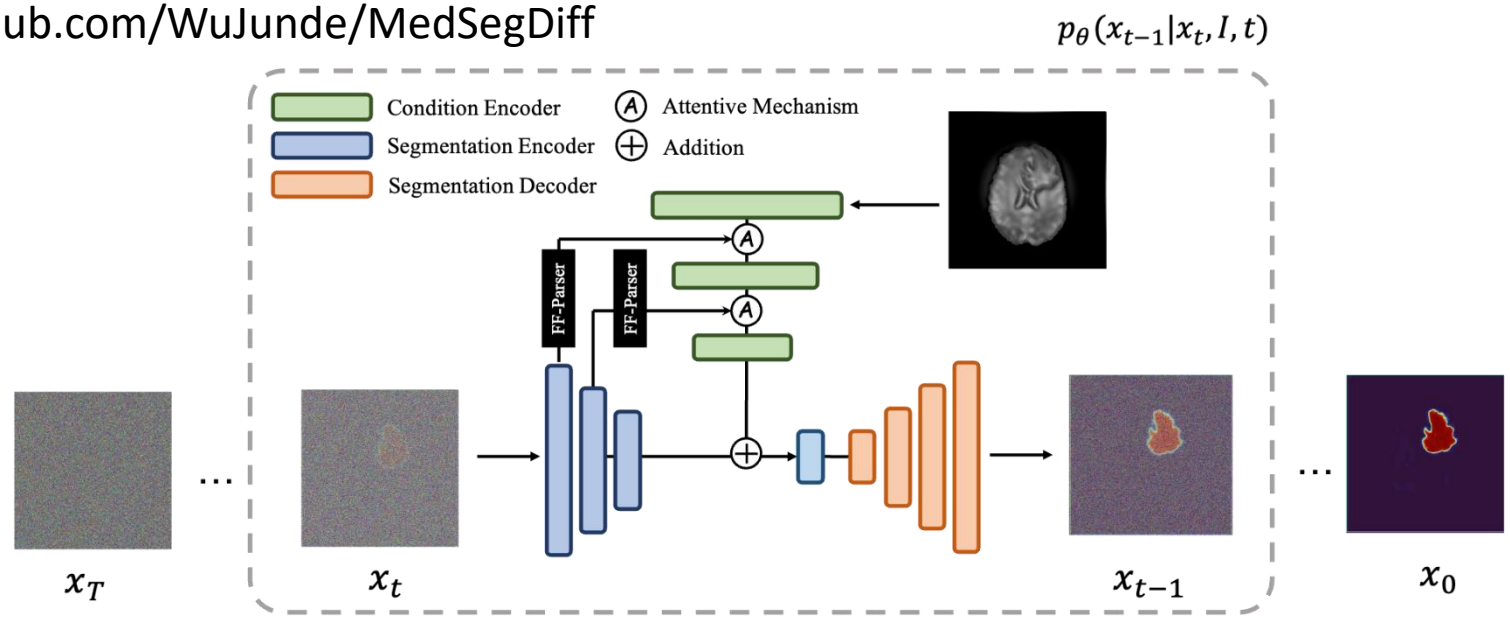
Latent feature estimation	IoU	Dice	Precision
Autoencoder	$0.399_{\pm 0.076}$	$0.566_{\pm 0.079}$	$0.621_{\pm 0.109}$
Diffusion (Ours)	$\mathbf{0.471}_{\pm 0.076}$	$\mathbf{0.636}_{\pm 0.072}$	$\mathbf{0.701}_{\pm 0.115}$

05. 总结和思考

提出 DARL 的扩散模型，用于自监督的血管分割任务。该模型利用扩散模块和生成模块，在不使用标签的情况下学习血管的表示，并通过可切换 SPADE 层生成合成血管造影图像和血管分割掩膜，更有效地学习血管的语义信息。尽管扩散模块的训练结合了其他损失函数，但推断不是迭代的，只在一步中完成，这使得该模型比现有的扩散模型更快速和独特。作者使用各种医学血管数据集验证了该模型的优越性，并指出该模型具有鲁棒性。

Diffusion Models 在医学图像分割的另一个应用：

https://github.com/WuJunde/MedSegDiff



谢谢！