CVPR Nashville JUNE 11-15, 2025

Noise-Consistent Siamese-Diffusion for Medical Image Synthesis and Segmentation

Kunpeng Qiu, Zhiqiang Gao*, Zhiying Zhou, Mingjie Sun*, Yongxin Guo*

kunpeng_qiu@u.nus.edu

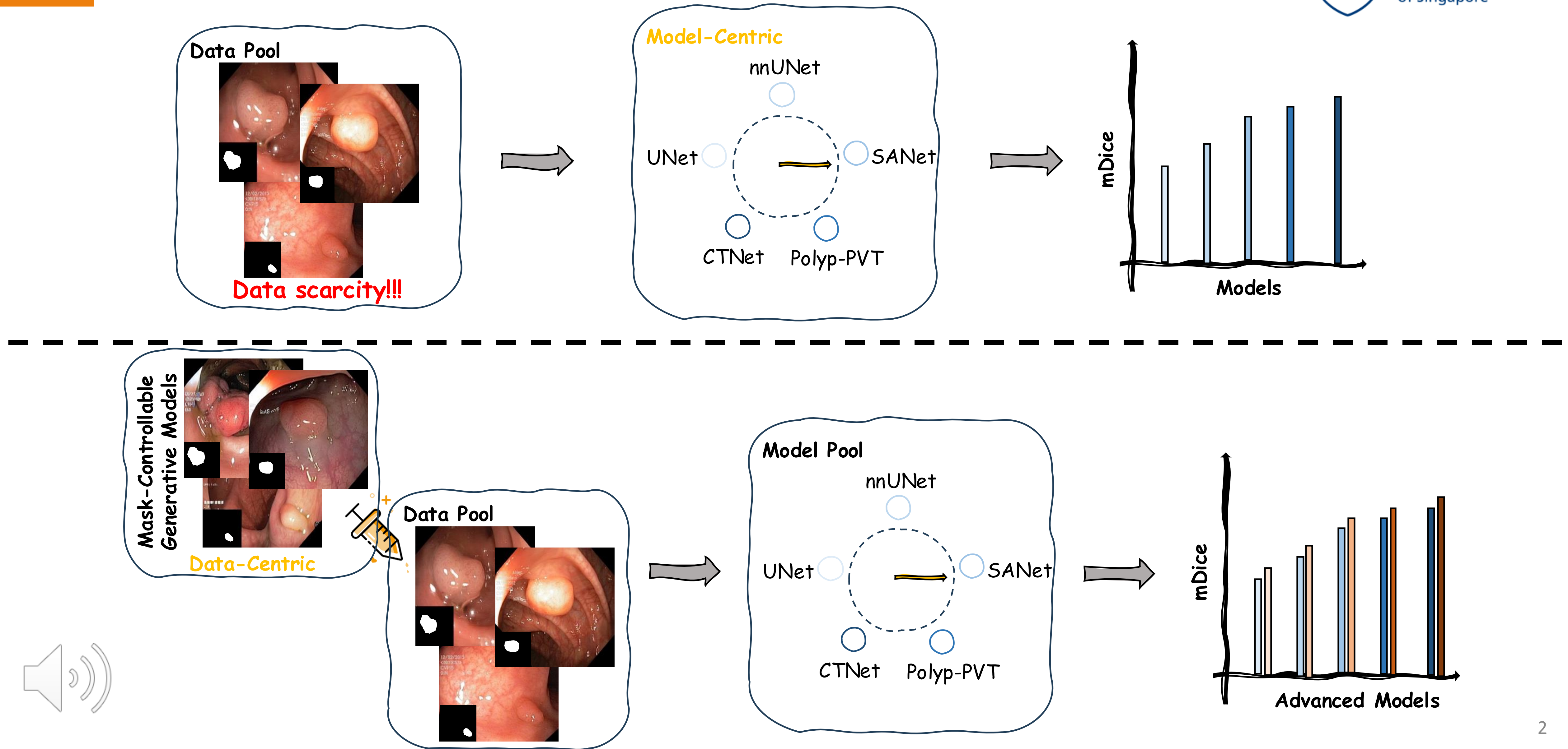
NUS National University of Singapore

苏州大学 SOOCHOW UNIVERSITY

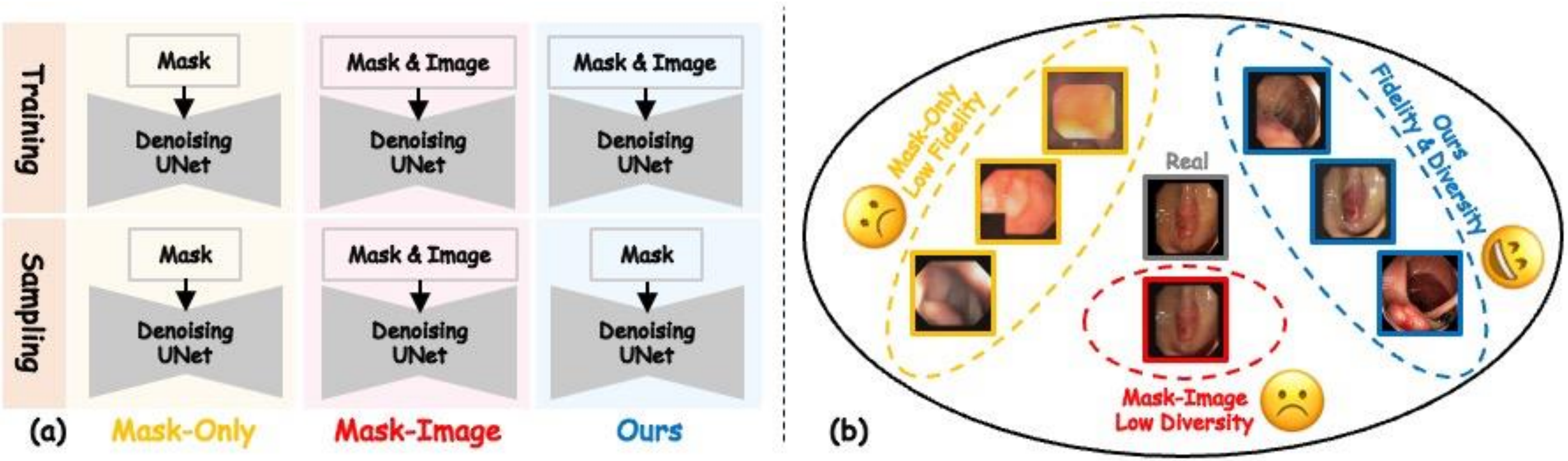
NUS National University of Singapore | 新加坡国立大学苏州研究院 NUS (Suzhou) Research Institute

CityU 香港城市大學 City University of Hong Kong

Introduction



Paradox

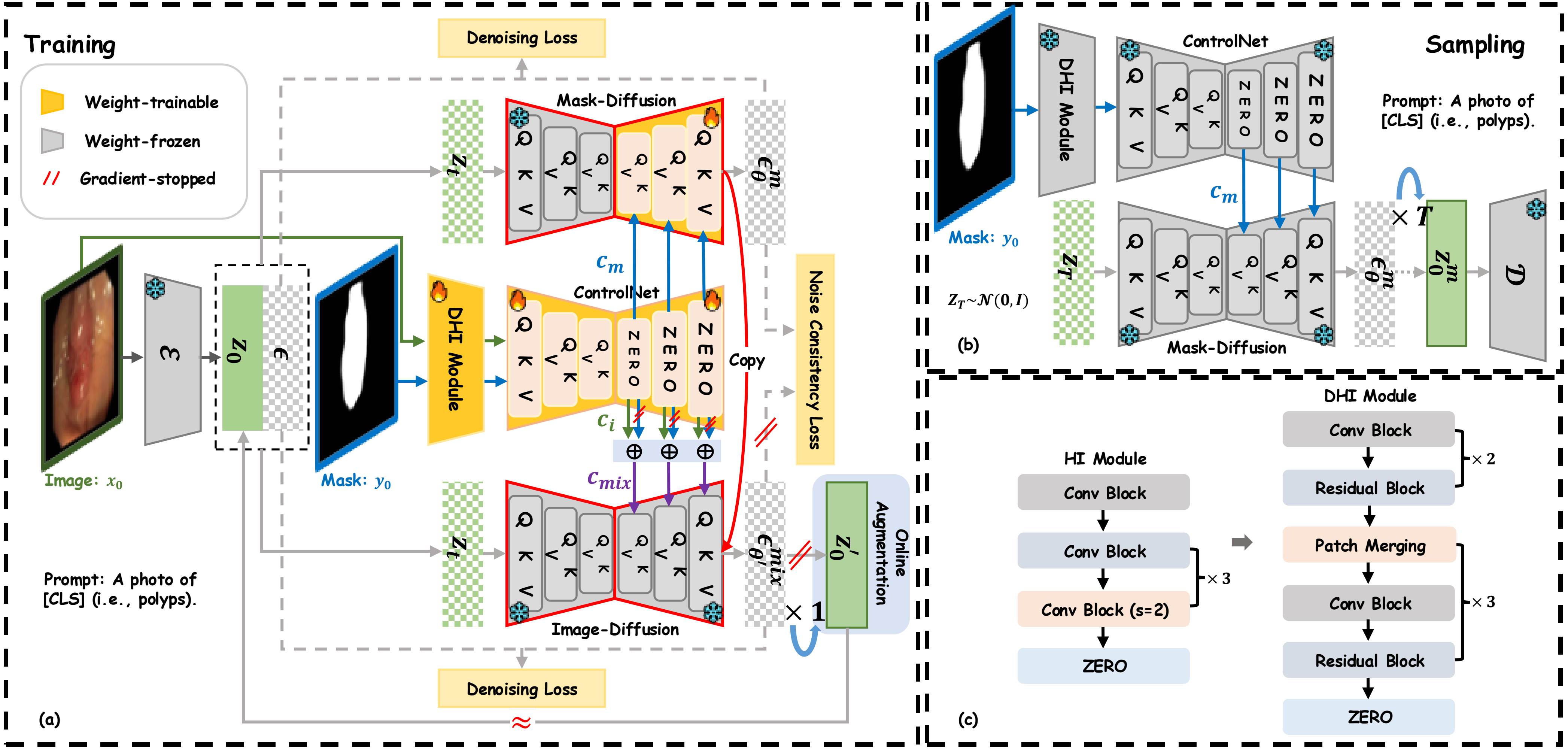


Training
Sampling
Mask
Denoising UNet
Mask & Image
Denoising UNet
Mask & Image
Denoising UNet
Mask
Denoising UNet
Mask & Image
Denoising UNet
Mask
Denoising UNet
(a)
Mask-Only
Mask-Image
Ours
(b)
Mask-Only Low Fidelity
Real
Ours Fidelity & Diversity
Mask-Image Low Diversity

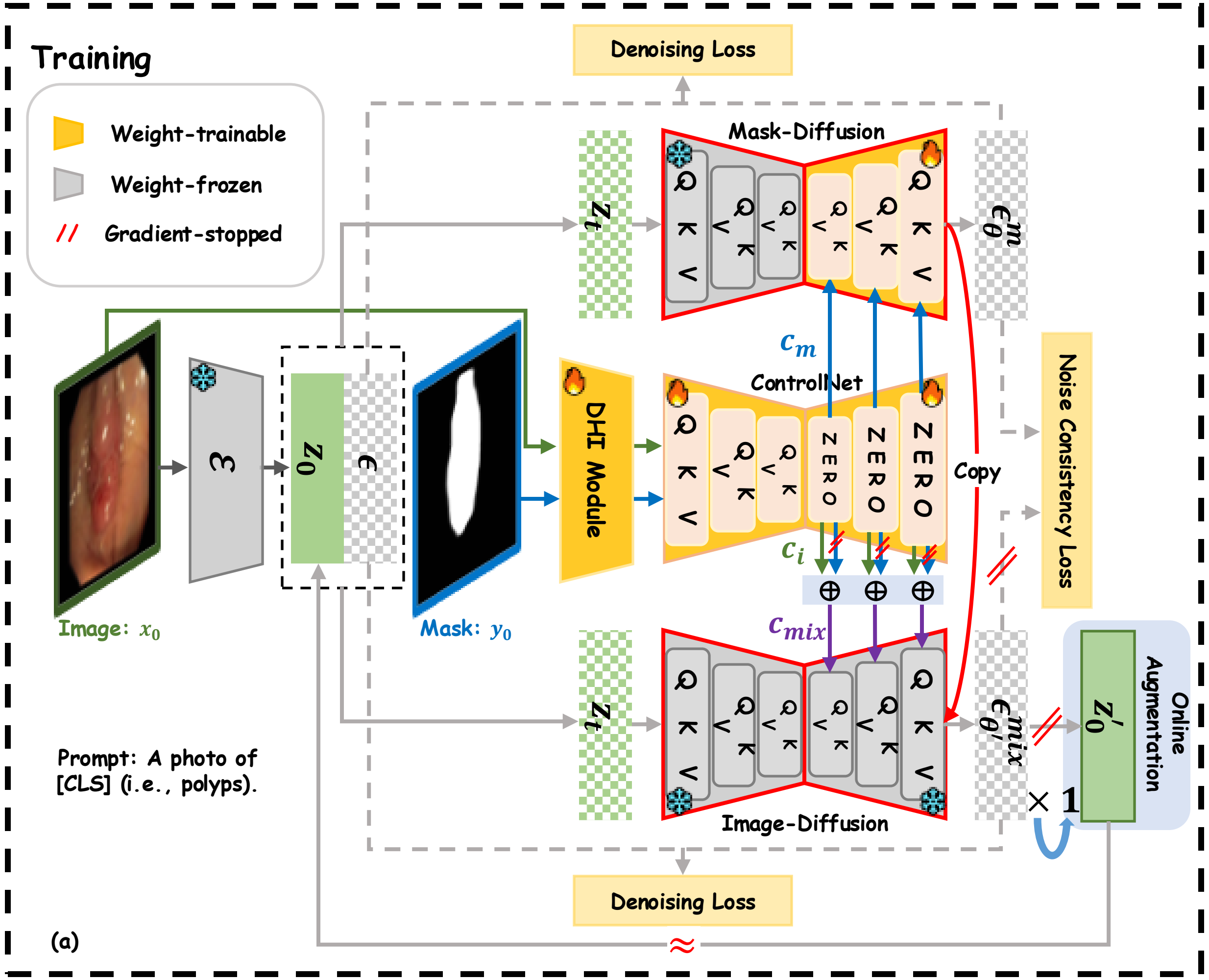
Main Contributions

- We argue that only synthesized medical images containing essential morphological characteristics can ensure the reliability of enhanced segmentation models.
- We propose the Siamese-Diffusion model to generate medical images with realistic morphological features.
- Extensive experiments demonstrate that our method outperforms existing approaches in both image quality and segmentation performance.

Overall Methodology



Noise Consistency Loss

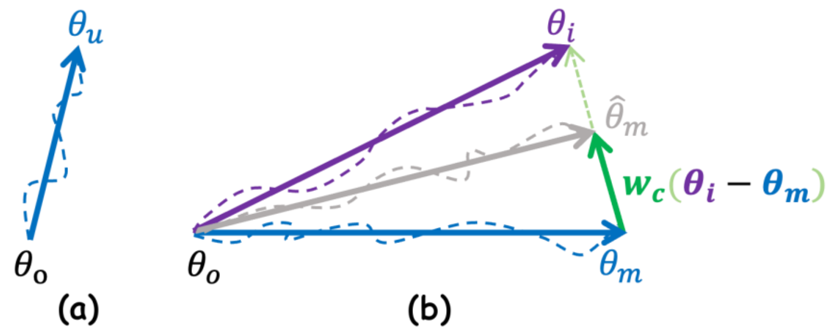


Denoising Loss:

$$c_{mix} = w_i \cdot c_i + w_m \cdot sg[c_m]$$

$$\mathcal{L}_m = \mathbb{E}\left[||\epsilon_\theta^m - \epsilon||^2\right], \epsilon_\theta^m = \epsilon_\theta(z_t, t, c_t, c_m)$$

$$\mathcal{L}_i = \mathbb{E}\left[||\epsilon_{\theta'}^{mix} - \epsilon||^2\right], \epsilon_{\theta'}^{mix} = \epsilon_{\theta'}(z_t, t, c_t, c_{mix})$$

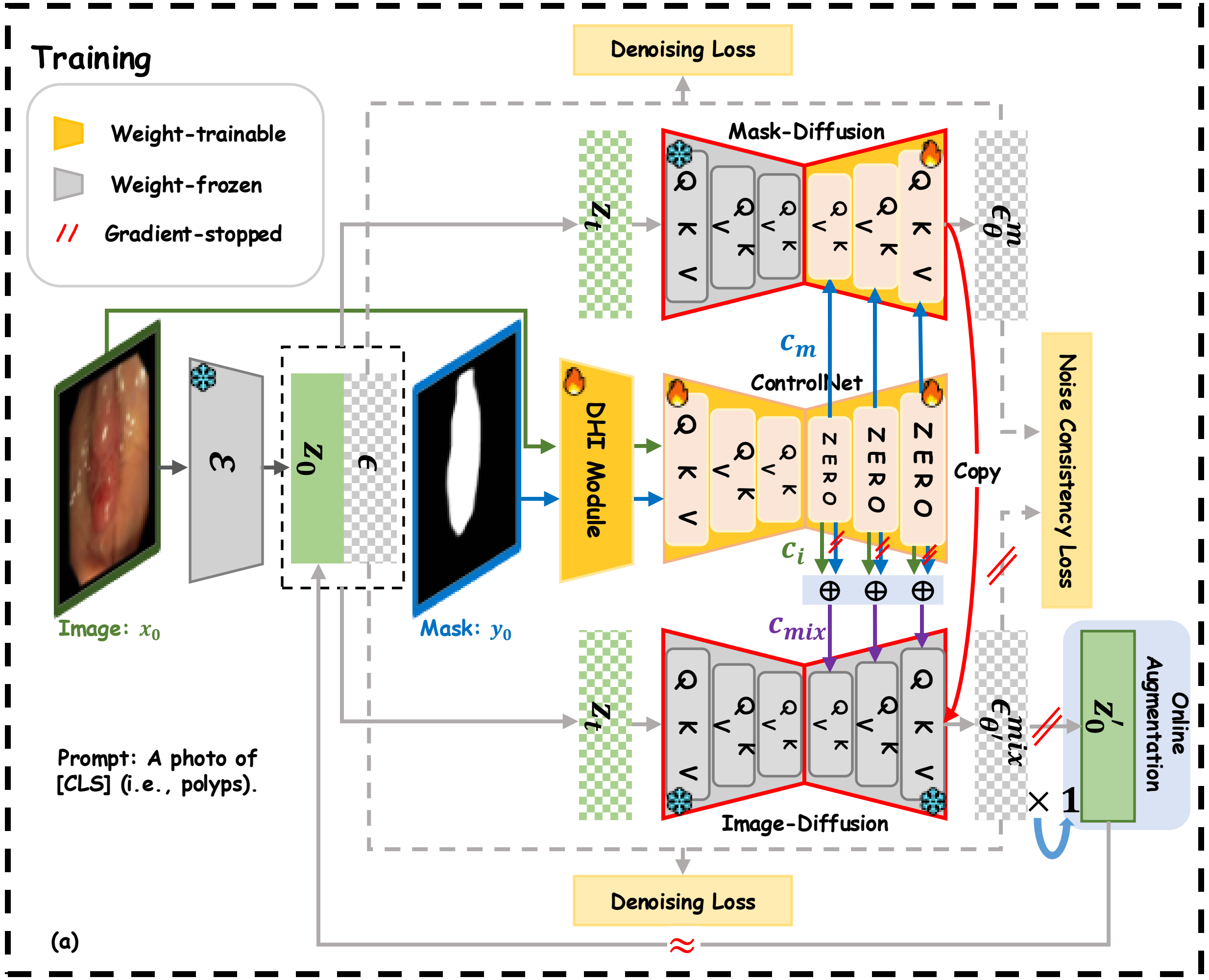


Consistency Loss:

$$\mathcal{L}_c = w_c \cdot \mathbb{E}\left[||\epsilon_\theta^m - sg[\epsilon_{\theta'}^{mix}]||^2\right]$$

$$\hat{\theta}_m = \theta_m + w_c \cdot (\theta_i - \theta_m)$$

Online-Augmentation



Online-Aug Loss:

$$x_t = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

$$z_0 \approx z_0' = \frac{z_t - \sqrt{1-\bar{\alpha}_t} sg[\epsilon_{\theta'}^{mix}]}{\sqrt{\bar{\alpha}_t}}$$

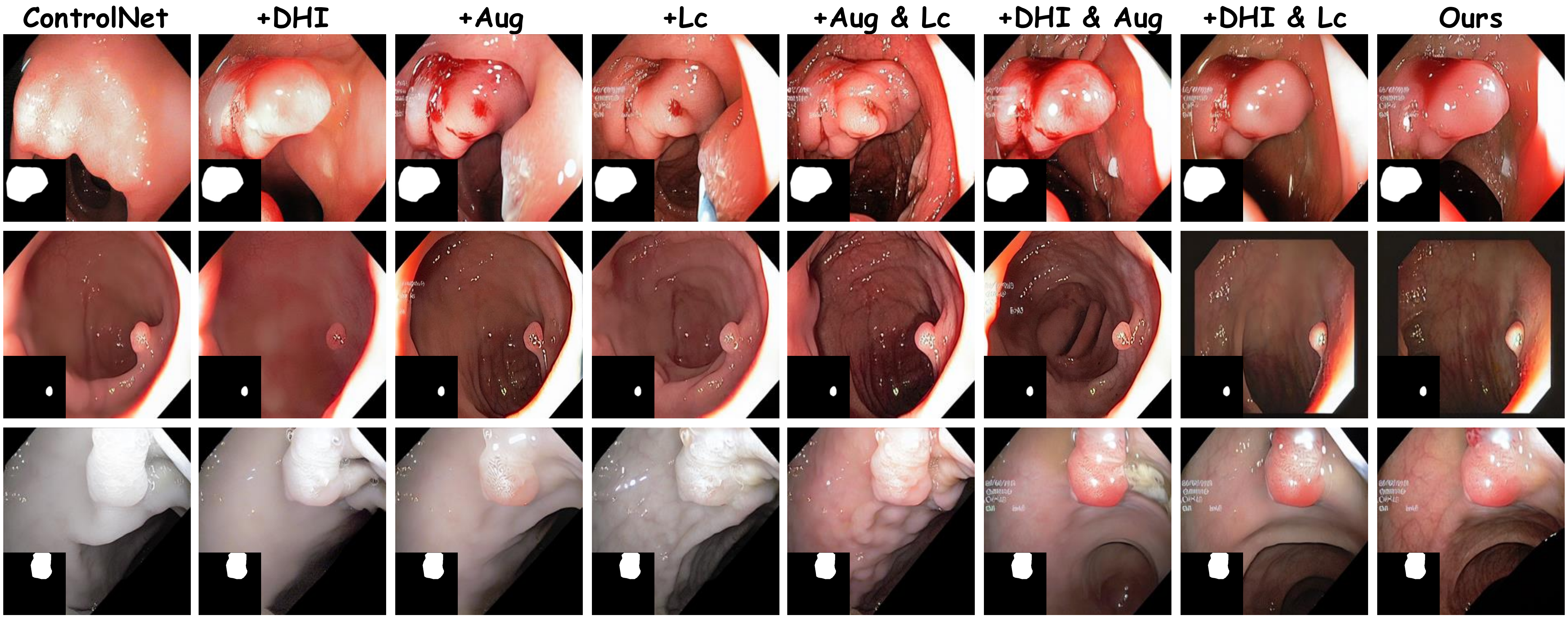
$$\mathcal{L}_{m'} = w_a \cdot \mathbb{E}\left[||\epsilon_\theta^{m'} - \epsilon||^2\right], \epsilon_\theta^{m'} = \epsilon_\theta(z_t', t, c_t, c_m)$$

$$w_a = \begin{cases} 1, & \text{if } k > K_\tau \text{ and } t < T_\tau, \\ 0, & \text{otherwise,} \end{cases}$$

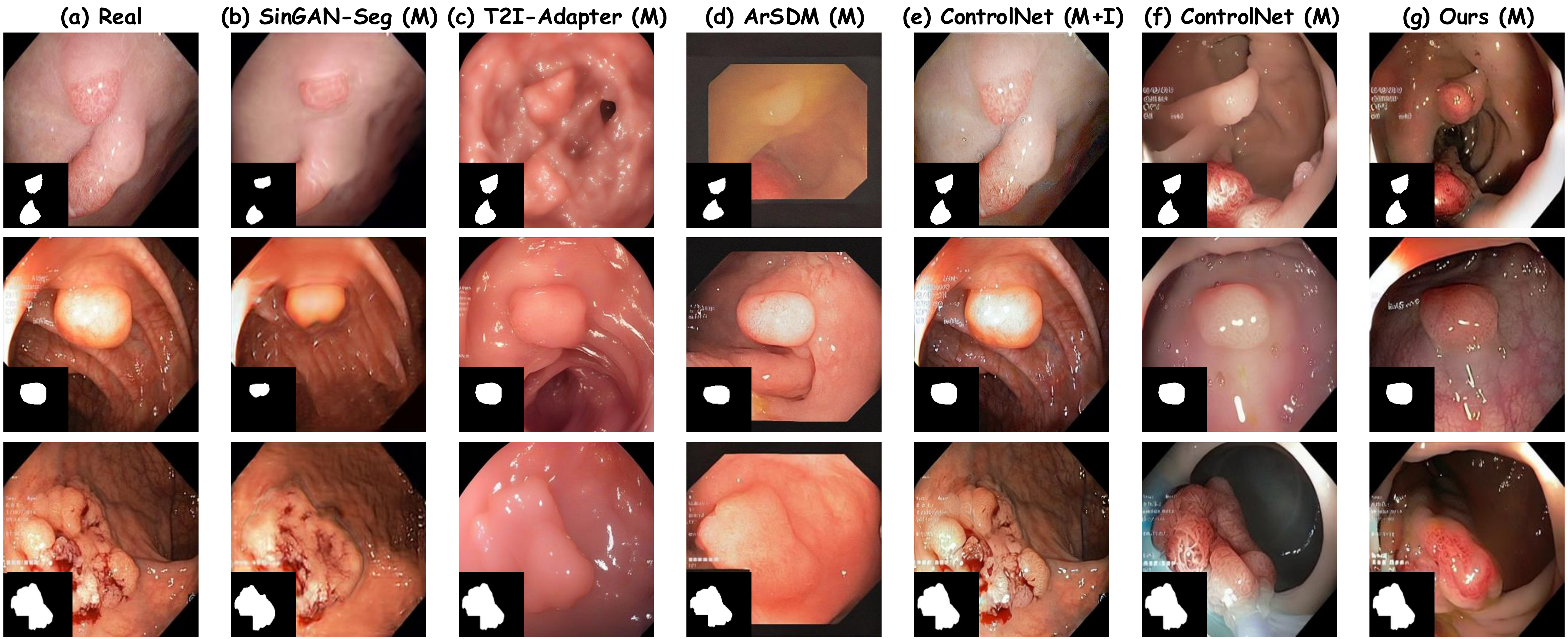
Overall Loss:

$$\mathcal{L} = \mathcal{L}_m + \mathcal{L}_i + \mathcal{L}_c + \mathcal{L}_{m'}$$

Ablation Study



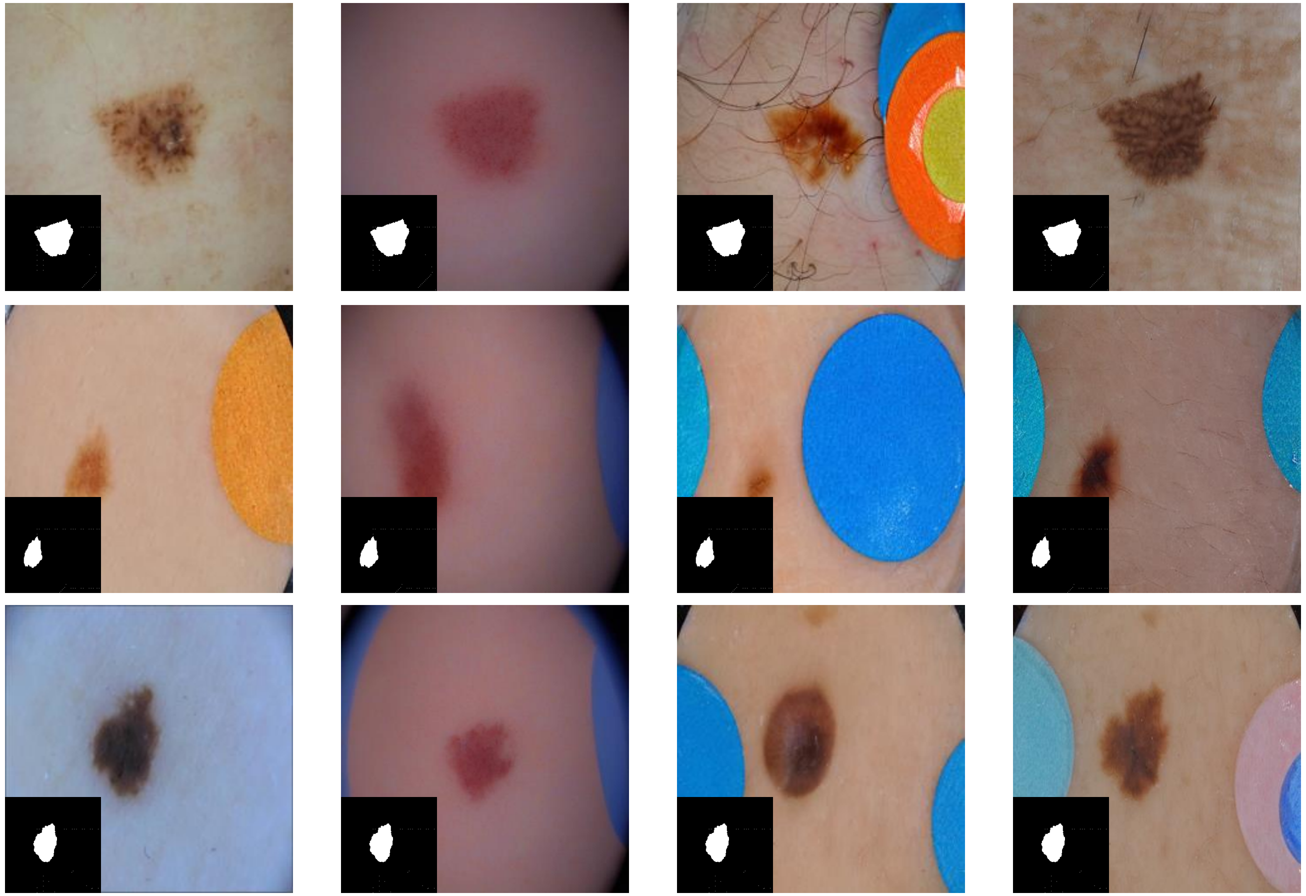
Method Comparison



Method Comparison

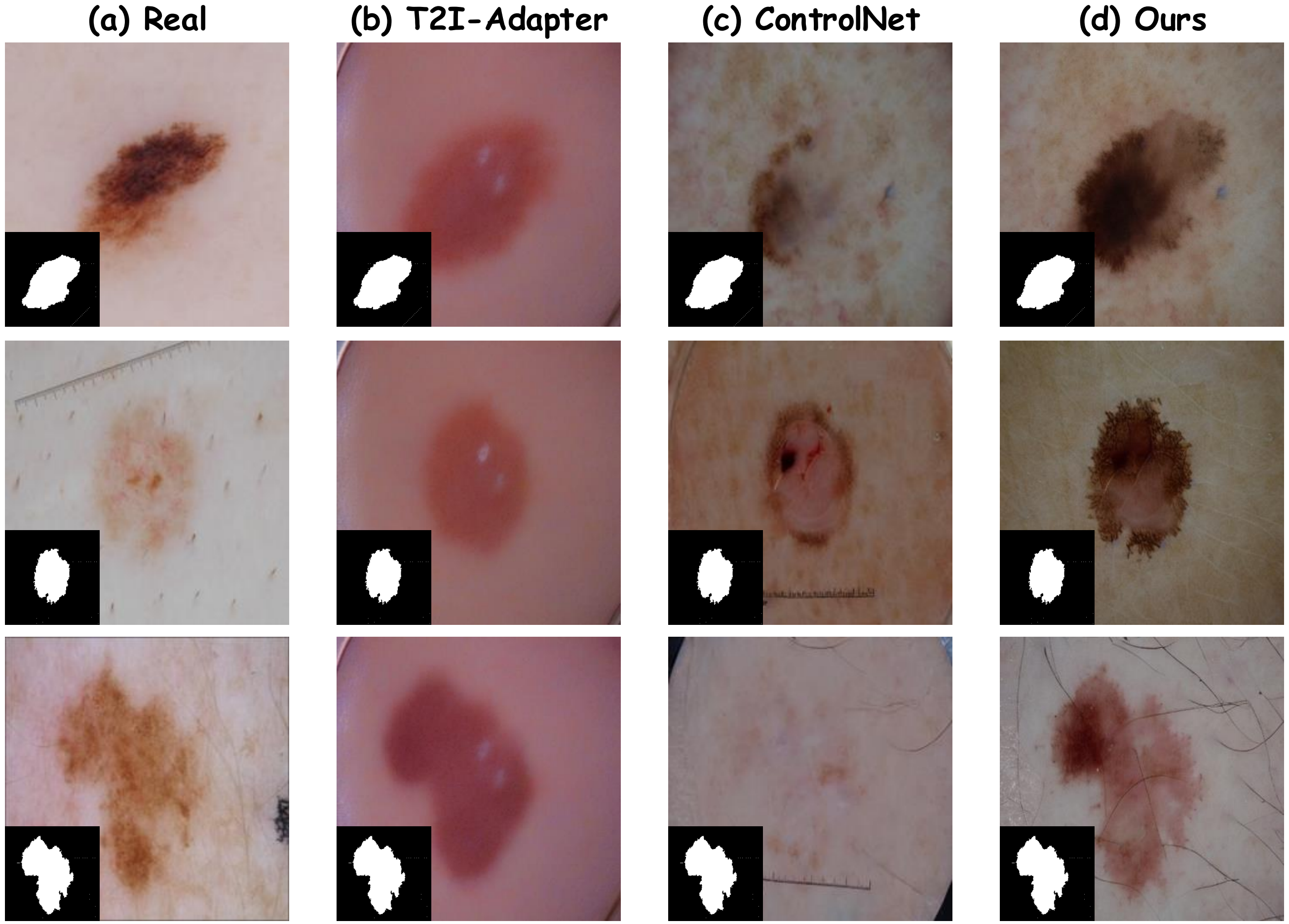
Methods	EndoScene [51]		ClinicDB [2]		Kvasir [26]		ColonDB [46]		ETIS [42]		Overall	
	mDice	mIoU	mDice	mIoU	mDice	mIoU	mDice	mIoU	mDice	mIoU	mDice	mIoU
SANet [52]	88.8	81.5	91.6	85.9	90.4	84.7	75.3	67.0	75.0	65.4	79.4	71.4
+Copy-Paste	89.7	83.0	90.2	85.1	90.3	84.8	77.7	70.0	77.4	68.8	81.1	73.7
+SinGAN-Seg [47] (M)	88.3	81.6	90.9	85.3	91.0	**85.8**	77.3	69.4	73.7	65.4	80.0	72.6
+T2I-Adapter [31] (M)	86.5	78.9	90.6	85.0	89.9	84.1	77.7	69.4	78.6	69.7	81.1	73.2
+ArSDM [14] (M)	90.2	83.2	91.4	86.1	91.1	85.6	77.7	70.0	78.0	69.5	81.5	74.1
+ControlNet [59] (M)	90.2	83.7	91.6	85.4	90.6	85.0	77.0	69.0	75.8	66.5	80.5	72.8
+ControlNet [59] (M+I)	89.9	83.3	91.4	86.2	**91.2**	85.5	78.2	70.5	75.4	66.6	81.0	73.6
+Ours (M)	**90.4**	**83.9**	**93.0**	**88.1**	**91.2**	85.5	**79.1**	**71.3**	**81.1**	**73.4**	**83.0**	**75.8**
Polyp-PVT [10]	90.0	83.3	93.7	88.9	91.7	86.4	80.8	72.7	78.7	70.6	83.3	76.0
+Copy-Paste	88.0	80.9	93.4	88.7	91.7	87.1	79.8	71.8	79.2	71.3	82.8	75.6
+SinGAN-Seg [47] (M)	87.0	79.7	91.7	87.0	**92.8**	**88.1**	76.9	69.0	74.2	66.7	80.1	73.0
+T2I-Adapter [31] (M)	88.7	81.6	93.1	88.1	91.3	86.0	80.6	72.4	79.0	71.0	83.1	75.7
+ArSDM [14] (M)	88.2	81.2	92.2	87.5	91.5	86.3	**81.7**	73.8	80.6	72.9	84.0	76.7
+ControlNet [59] (M)	88.0	81.1	93.7	89.0	91.5	86.3	80.8	72.5	76.1	68.1	82.5	75.1
+ControlNet [59] (M+I)	90.3	83.5	92.2	87.7	91.1	86.2	81.6	73.7	77.2	69.1	83.2	76.0
+Ours (M)	**90.4**	**83.8**	**93.9**	**89.3**	91.1	85.9	**81.7**	**74.0**	**81.5**	**73.4**	**84.4**	**77.3**
CTNet [54]	90.8	84.4	93.6	88.7	91.7	86.3	81.3	73.4	**81.0**	73.4	84.2	77.0
+Copy-Paste	**91.6**	**85.5**	92.3	87.1	91.5	86.3	82.5	74.7	78.1	70.6	84.0	76.9
+SinGAN-Seg [47] (M)	89.4	82.3	93.6	88.8	**92.9**	88.6	78.0	70.4	76.4	68.7	81.5	74.6
+T2I-Adapter [31] (M)	90.8	84.3	93.7	88.7	91.9	86.9	81.7	74.1	80.4	72.5	84.3	77.2
+ArSDM [14] (M)	89.8	82.9	92.9	87.8	90.5	85.2	81.9	74.0	80.9	**73.5**	84.2	77.0
+ControlNet [59] (M)	90.2	84.1	92.3	87.5	91.7	86.9	81.4	73.5	79.5	71.7	83.7	76.6
+ControlNet [59] (M+I)	90.7	84.2	93.2	88.6	92.0	87.2	81.1	73.6	80.9	73.1	84.1	77.1
+Ours (M)	90.1	83.6	**93.8**	**88.9**	**92.9**	**89.1**	**83.8**	**75.8**	79.5	72.0	**85.1**	**78.1**

Extended Evaluation (ISIC2016)



Methods	ISIC2016 [8]			
	UNet [37]		SegFormer [55]	
	mDice	mIoU	mDice	mIoU
Real Dataset	89.58	86.62	93.57	89.06
+Copy-Paste	89.43	86.48	93.53	88.85
+T2I-Adapter [31]	89.48	86.47	93.71	89.24
+ControlNet [59]	89.42	86.51	93.54	88.83
+**Ours** (Real)	89.91	87.01	94.14	89.76
+**Ours** (Random)	**90.06**	**87.25**	**94.31**	**90.01**

Extended Evaluation (ISIC2018)



Methods	ISIC2018 [7]			
	UNet [37]		SegFormer [55]	
	mDice	mIoU	mDice	mIoU
Real Dataset	82.19	78.45	91.35	86.10
+Copy-Paste	81.00	77.16	91.25	86.09
+T2I-Adapter [31]	81.16	77.04	91.19	85.73
+ControlNet [59]	81.76	77.94	91.52	86.19
+Ours (Real)	82.81	79.04	91.80	86.67
+Ours (Random)	**83.71**	**80.09**	**91.93**	**87.12**

Conclusion

- Introduce a novel image synthesis method-Siamese-Diffusion.
- Extensive experiments across five datasets demonstrate the effectiveness and superiority of our method.

Thanks for your attention!

https://github.com/Qiukunpeng/Siamese-Diffusion