



浙江大學
ZHEJIANG UNIVERSITY



Highlight



SCSA: A Plug-and-Play Semantic Continuous-Sparse Attention for Arbitrary Semantic Style Transfer

Chunnan Shang, ZhiZhong Wang, Hongwei Wang, Xiangming Meng

Zhejiang University

Project Page



Problem Definition

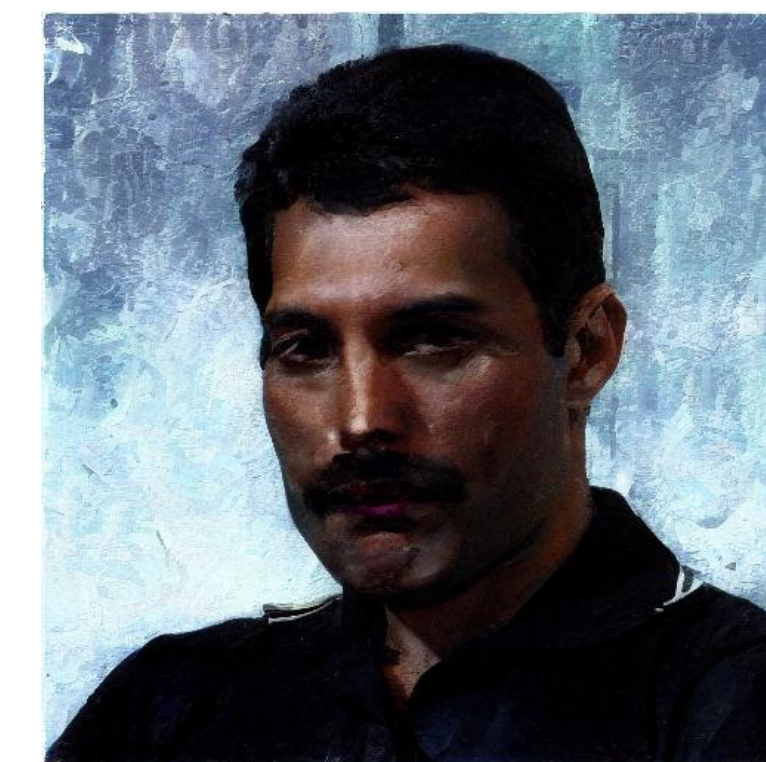
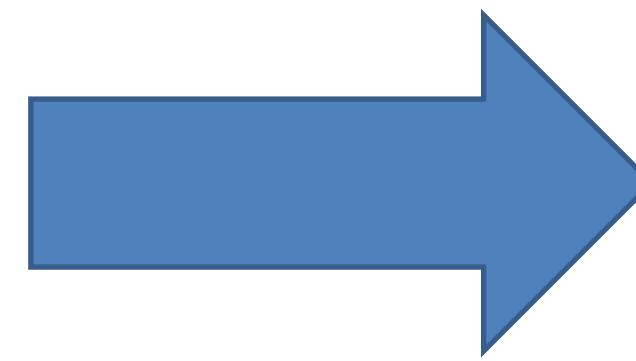
- Arbitrary style transfer aims to transfer the style of any given style image onto an arbitrary content image.



Content Image

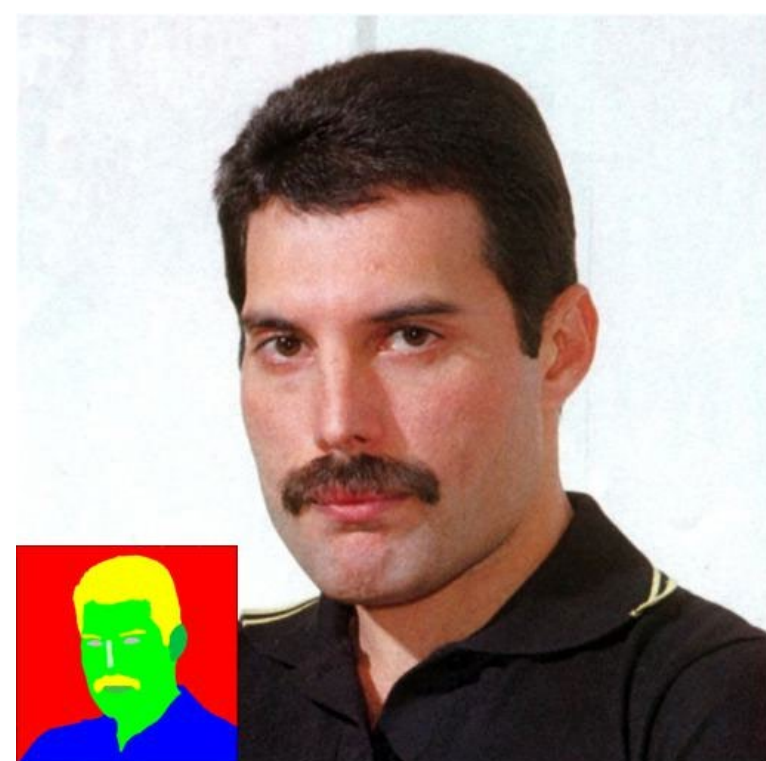


Style Image



Stylized Image

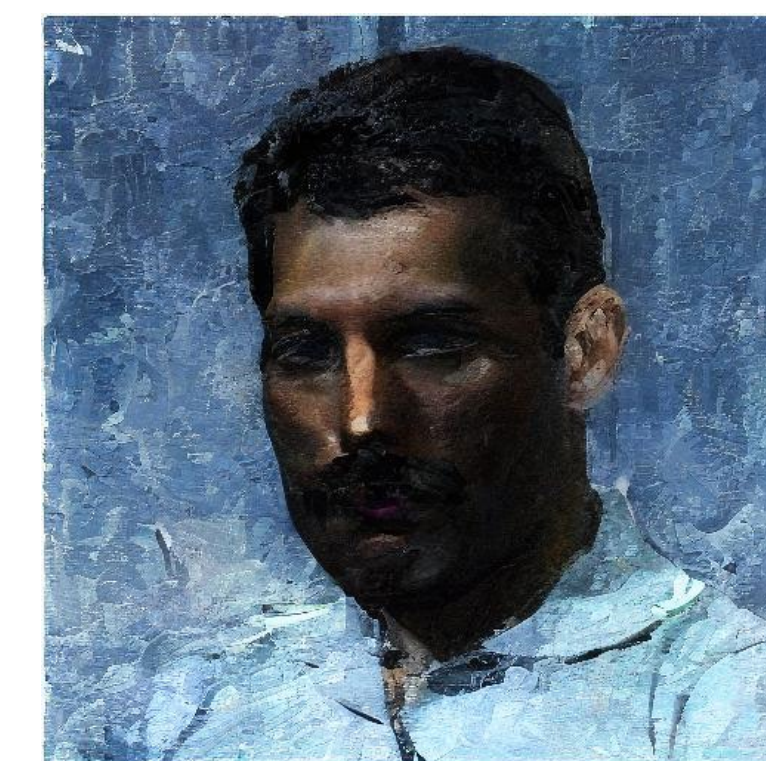
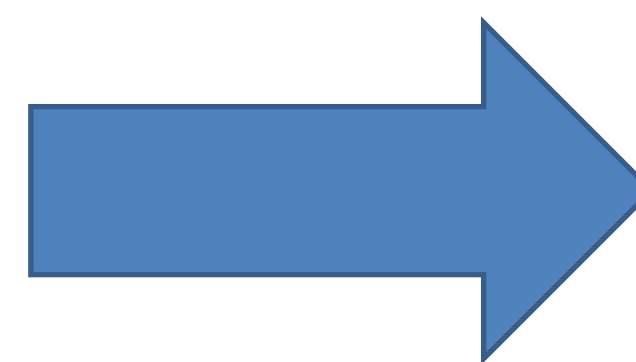
- Arbitrary semantic style transfer aims to transfer the style of the corresponding semantic region in any given style image onto the corresponding semantic region in an arbitrary content image guided by the corresponding semantic map.



Content Image



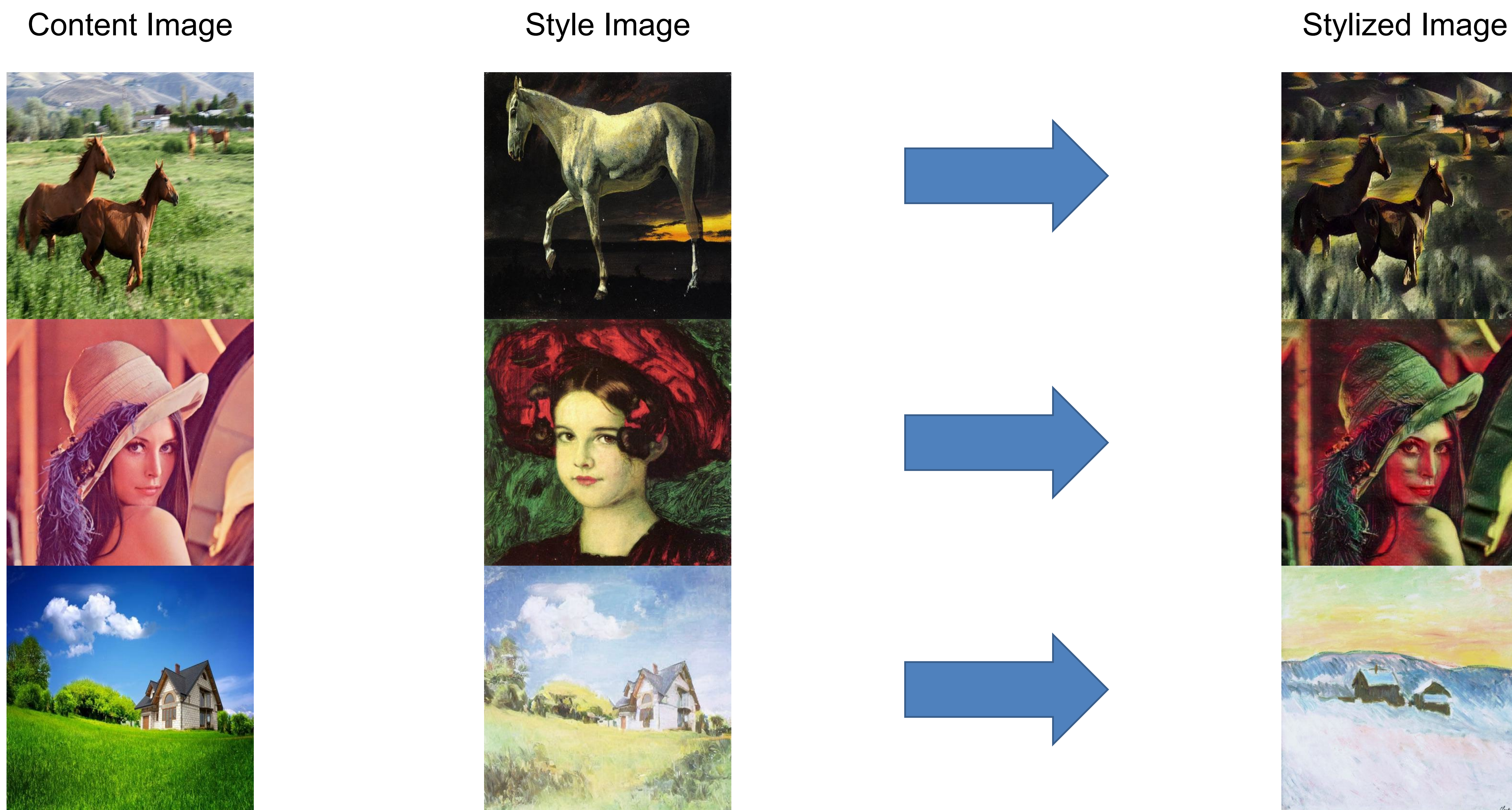
Style Image



Stylized Image

Background

- Attention-based arbitrary style transfer methods have flourished and produced high-quality stylized images. However, they perform poorly on the content and style images with the same semantics.



- We propose a plug-and-play semantic continuous sparse attention, dubbed SCSA, for arbitrary semantic style transfer.

Motivation

- StyleID [1] has achieved arbitrary style transfer by adjusting the injected features of the attention mechanism.
- Some studies [2] have shown that by adjusting the attention maps of the attention mechanism, the corresponding attributes of the generated images can be regulated.
- **Can we achieve semantic style transfer by regulating attention maps with the image information learned by the model?**

[1] Chung J, Hyun S, Heo J P. Style injection in diffusion: A training-free approach for adapting large-scale diffusion models for style transfer[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2024: 8795-8805.

[2] Feng Z, Zhang Z, Yu X, et al. Ernie-vilg 2.0: Improving text-to-image diffusion model with knowledge-enhanced mixture-of-denoising-experts[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 10135-10145.

Our Method — SCSA

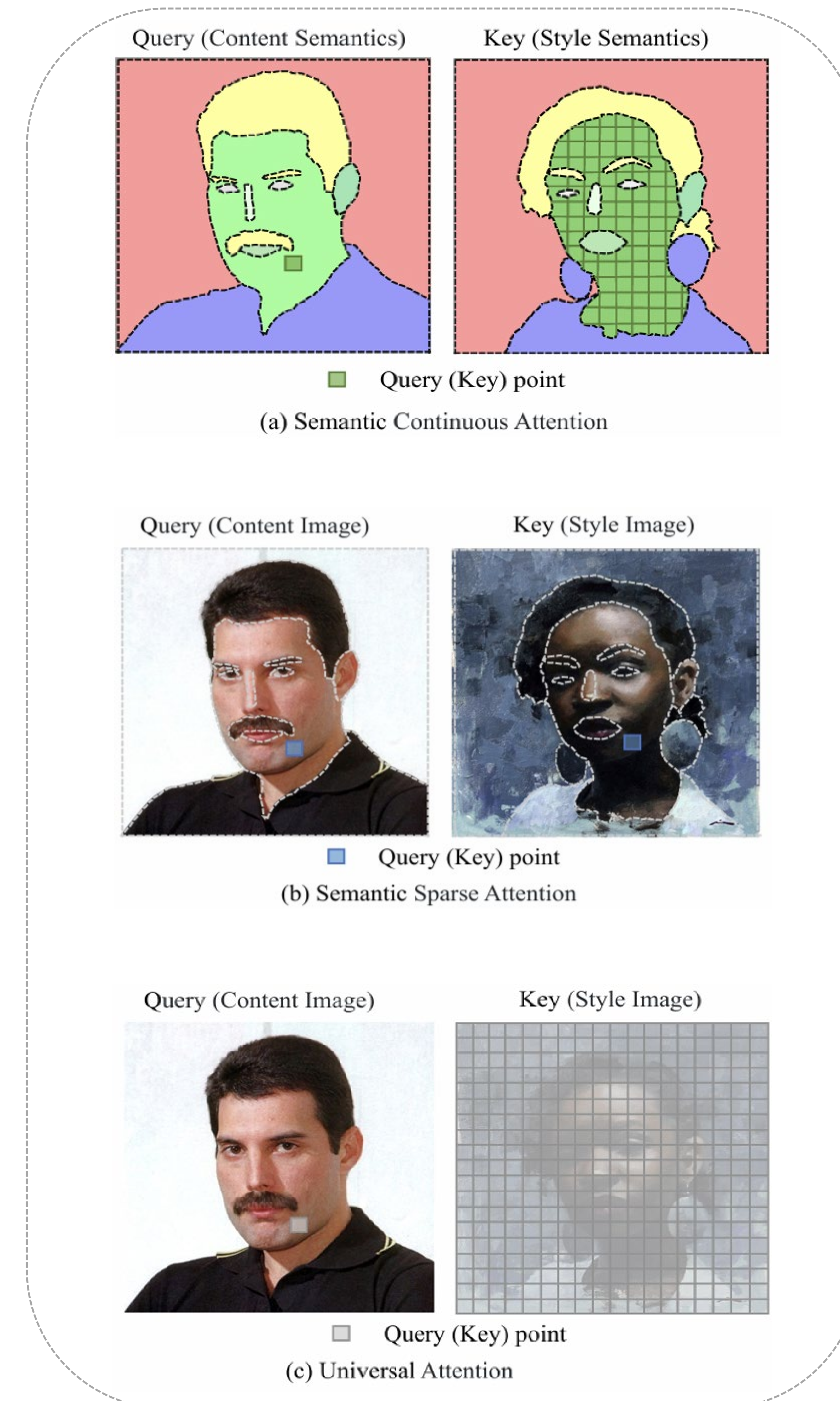
➤ Semantic Continuous-Sparse Attention (SCSA)

● Semantic Continuous Attention (SCA)

The query point of the content semantic map features can match all continuous key points of the style semantic map features in the same semantic region.

● Semantic Sparse Attention (SSA)

The query point of the content image features can match the most similar sparse key point of the style image features in the same semantic region.

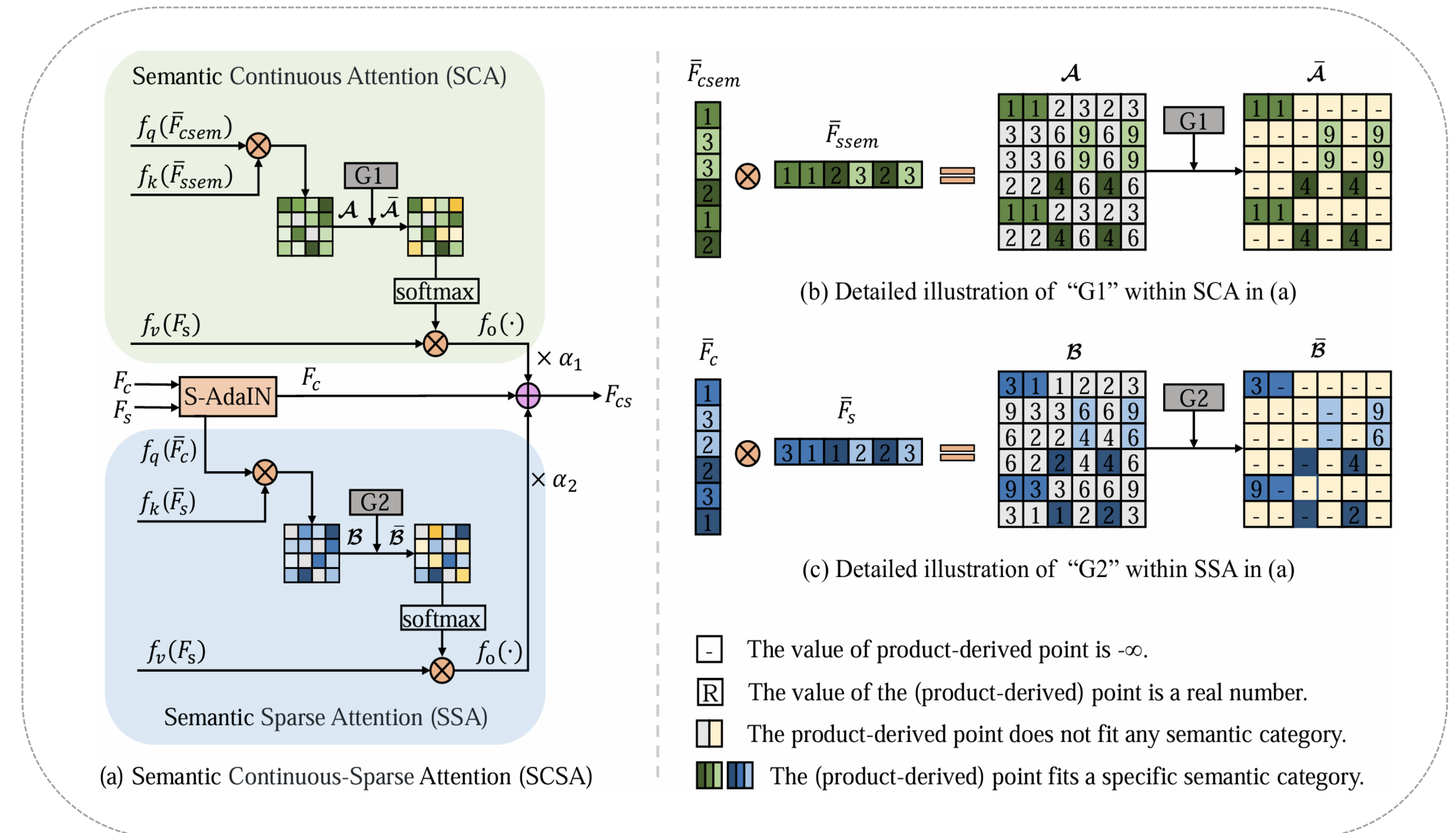


Our Method — SCSA

➤ Semantic Continuous-Sparse Attention (SCSA)

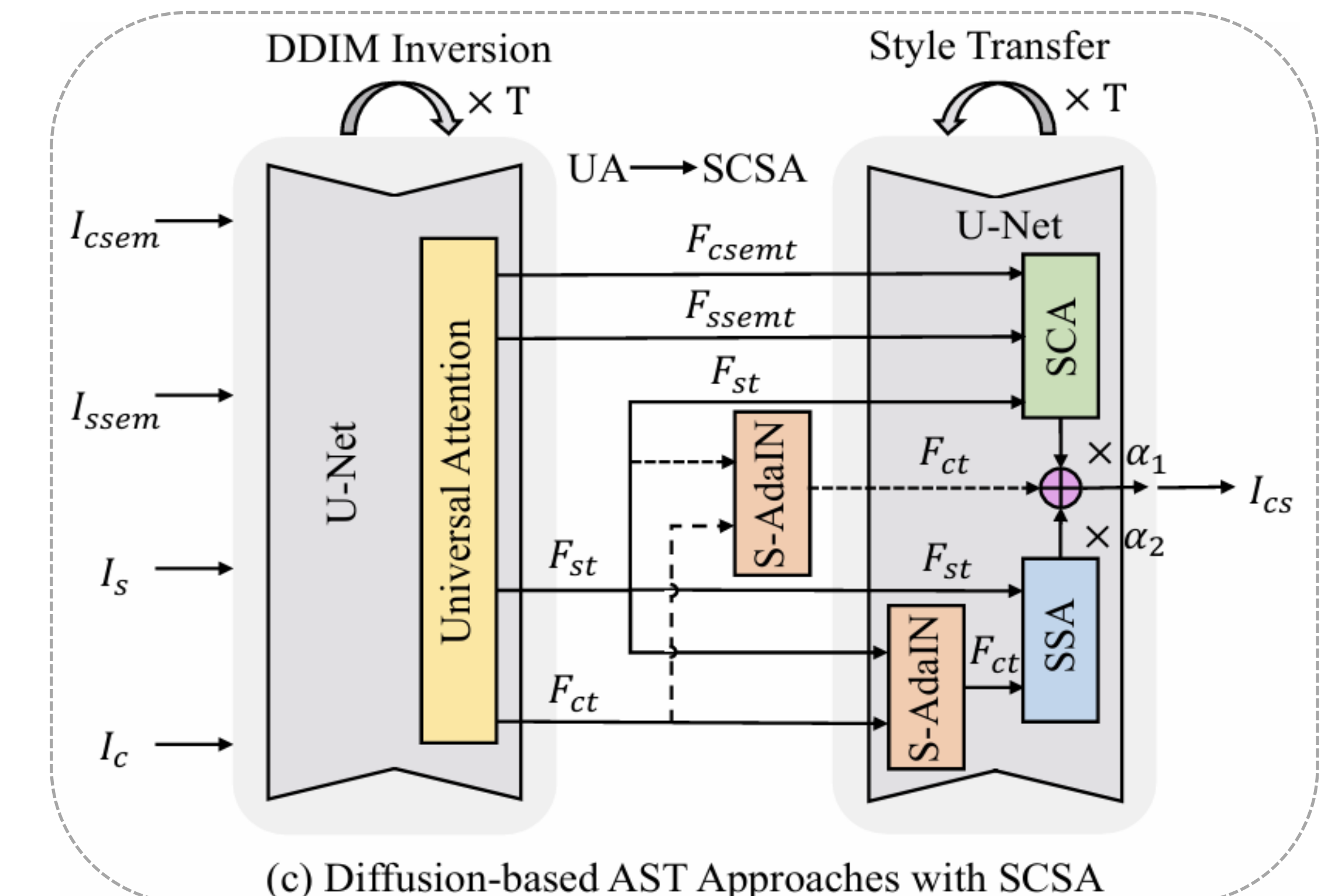
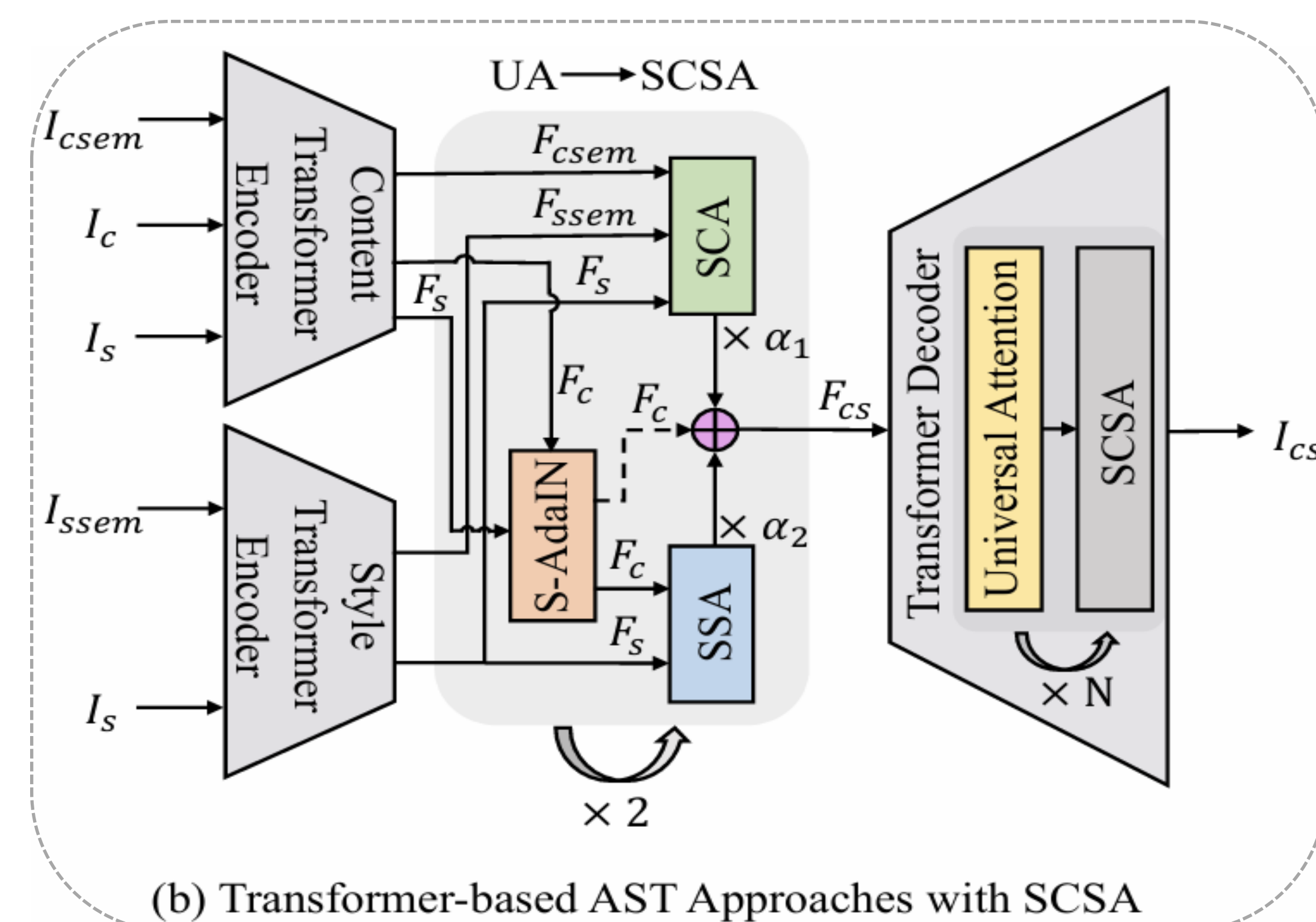
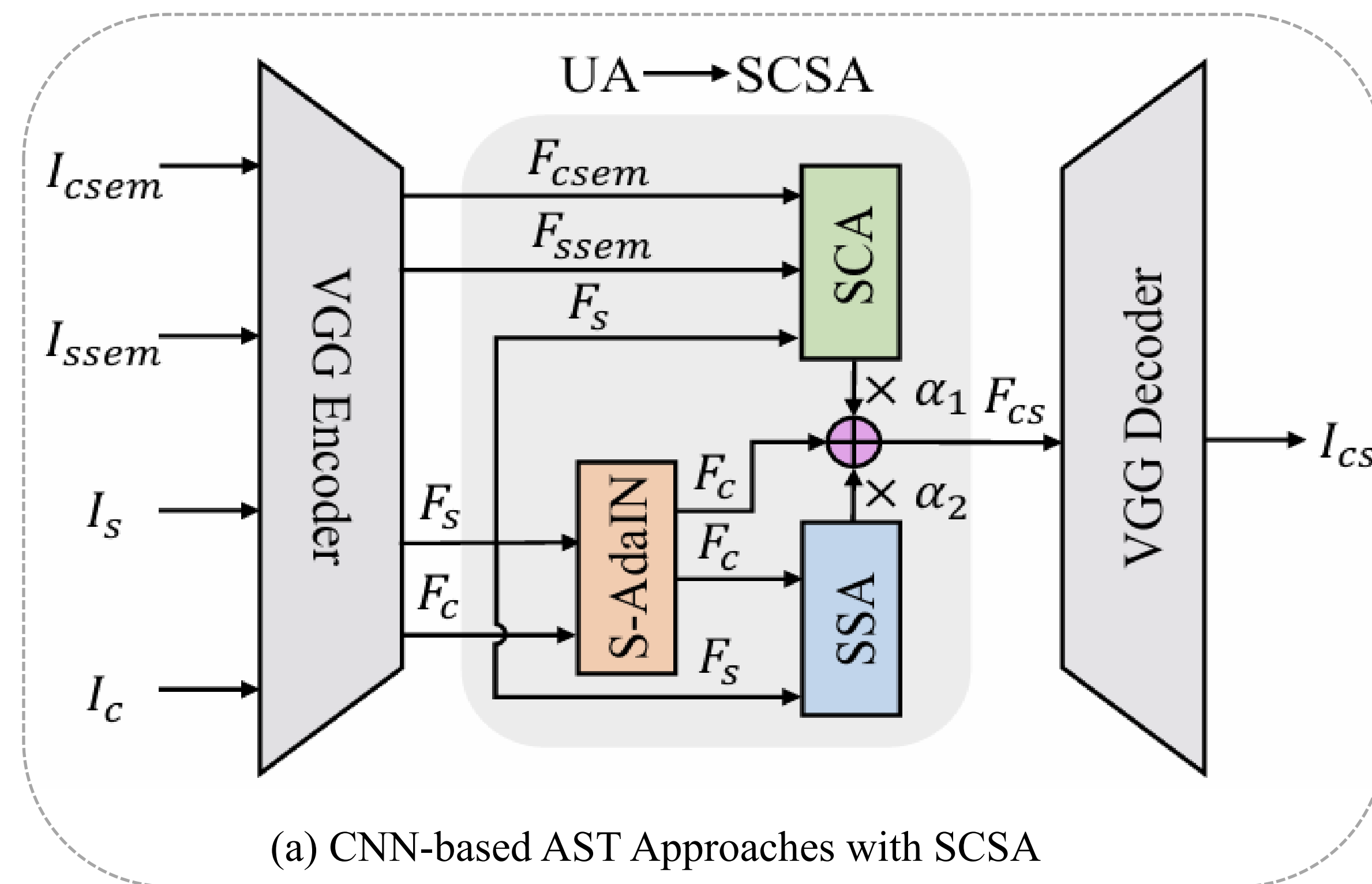
- Semantic Continuous Attention (SCA)

- Semantic Sparse Attention (SSA)



Embedding Our Method into Existing Frameworks

- CNN-based arbitrary style transfer methods with SCSA.
- Transformer-based arbitrary style transfer methods with SCSA.
- Diffusion-based arbitrary style transfer methods with SCSA.



Experimental Results

➤ Qualitative results

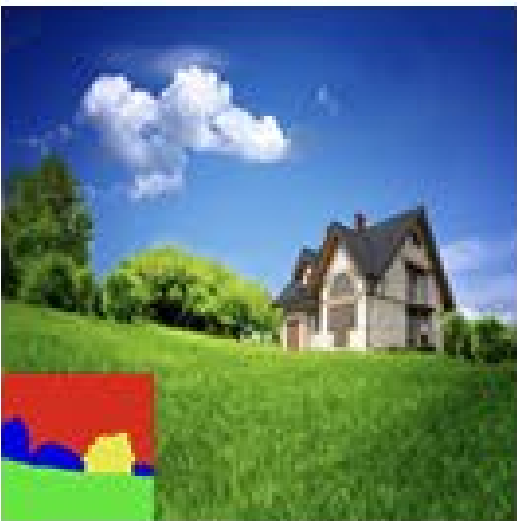
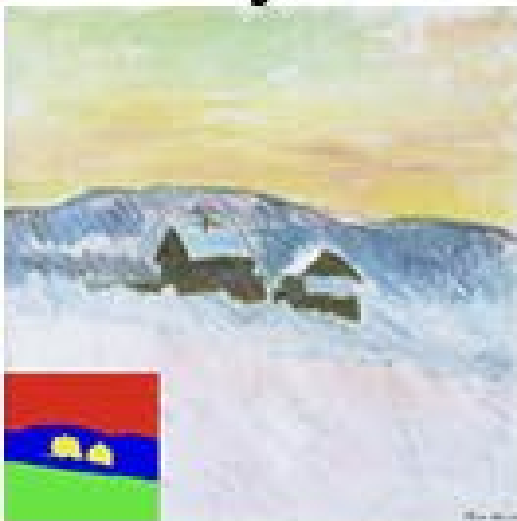
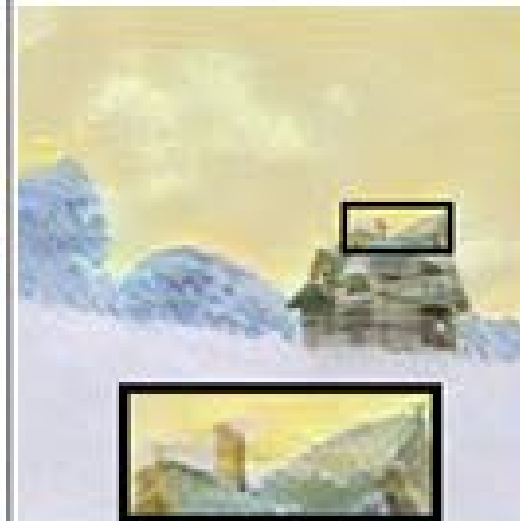
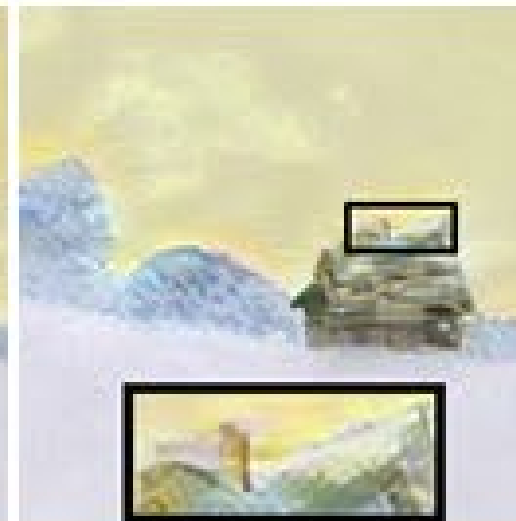
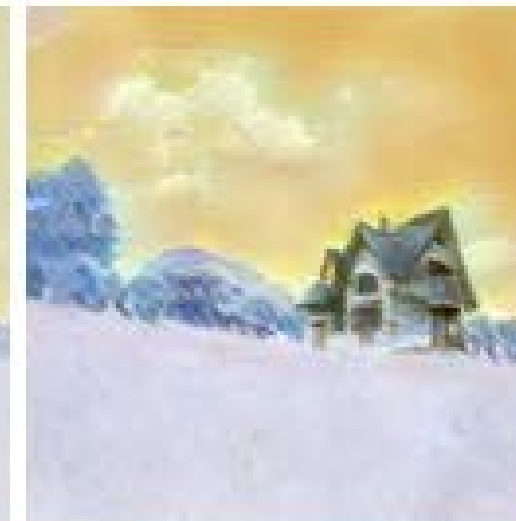

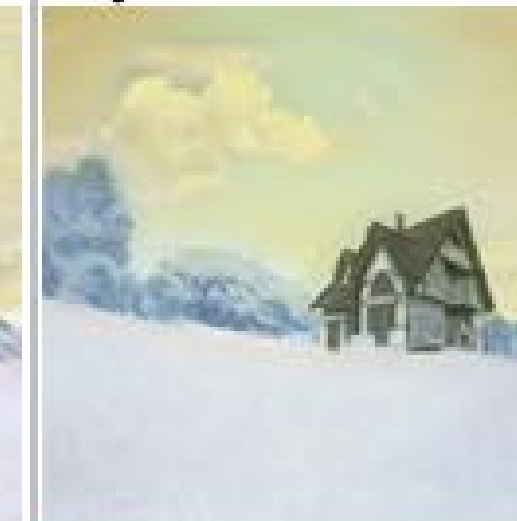

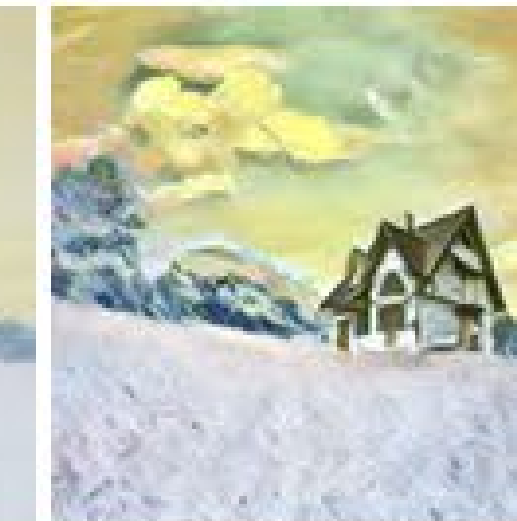



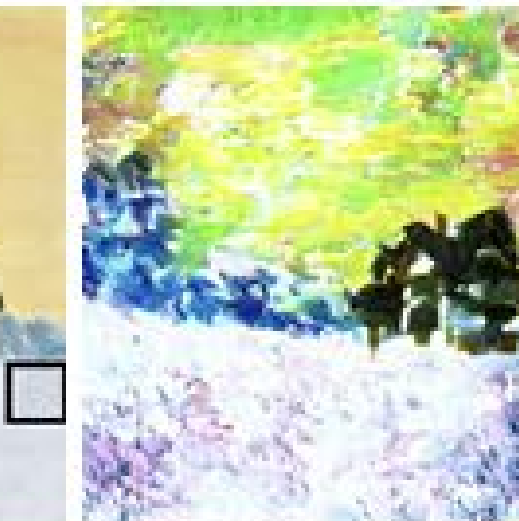





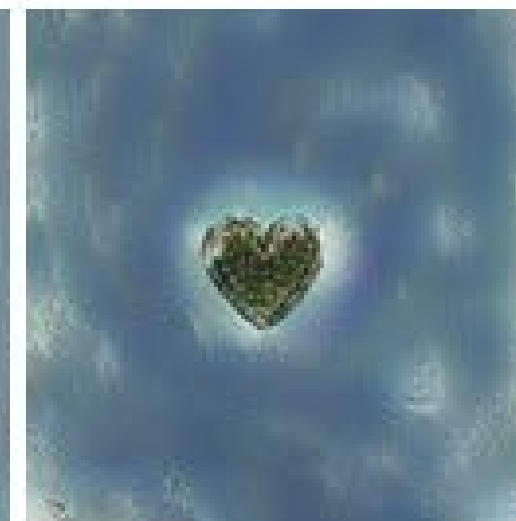

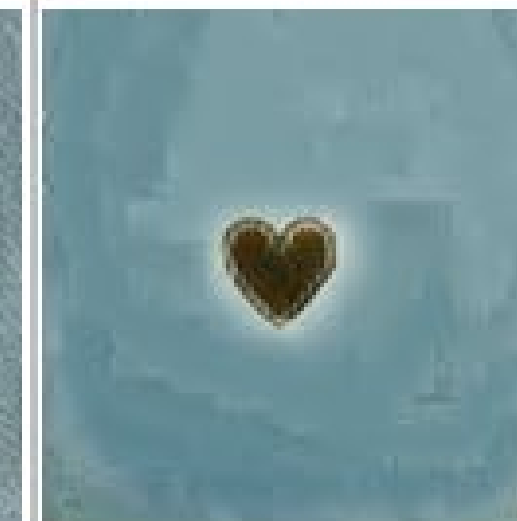





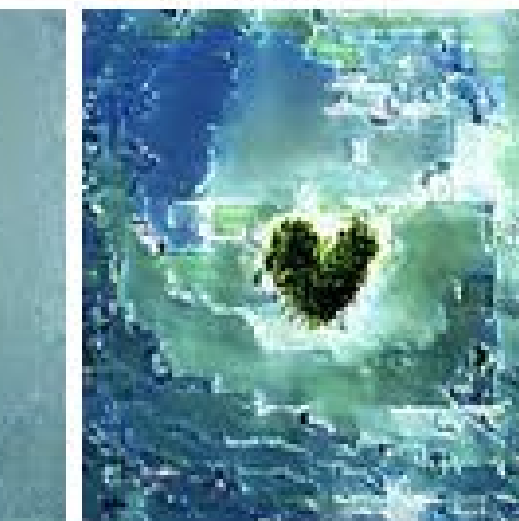
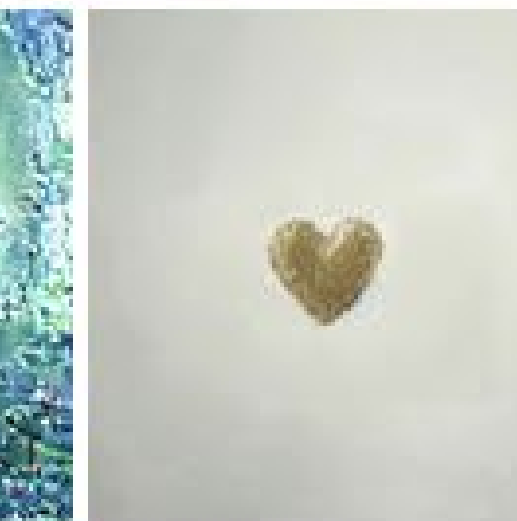


➤ Quantitative results

	SANet	SANet + SCSA	StyTR ²	StyTR ² + SCSA	StyleID	StyleID + SCSA	STROTSS	MAST	TR	DIA	GLStyleNet
SSL ↓	1.6583	0.8762	1.9826	1.2228	1.7538	1.2447	1.0981	1.7320	1.2631	1.9398	1.0305
FID ↓	14.3385	13.0788	12.5273	12.3963	12.5944	12.4497	12.8400	16.5163	13.5846	20.7942	14.4880
CFSD ↓	0.1103	0.0874	0.0752	0.0705	0.0916	0.1178	0.1008	0.0746	0.1149	0.1147	0.3991
Pref. ↑	0.1685	0.8315	0.1576	0.8424	0.2192	0.7808	0.1867	0.1100	0.0450	0.0167	0.1183

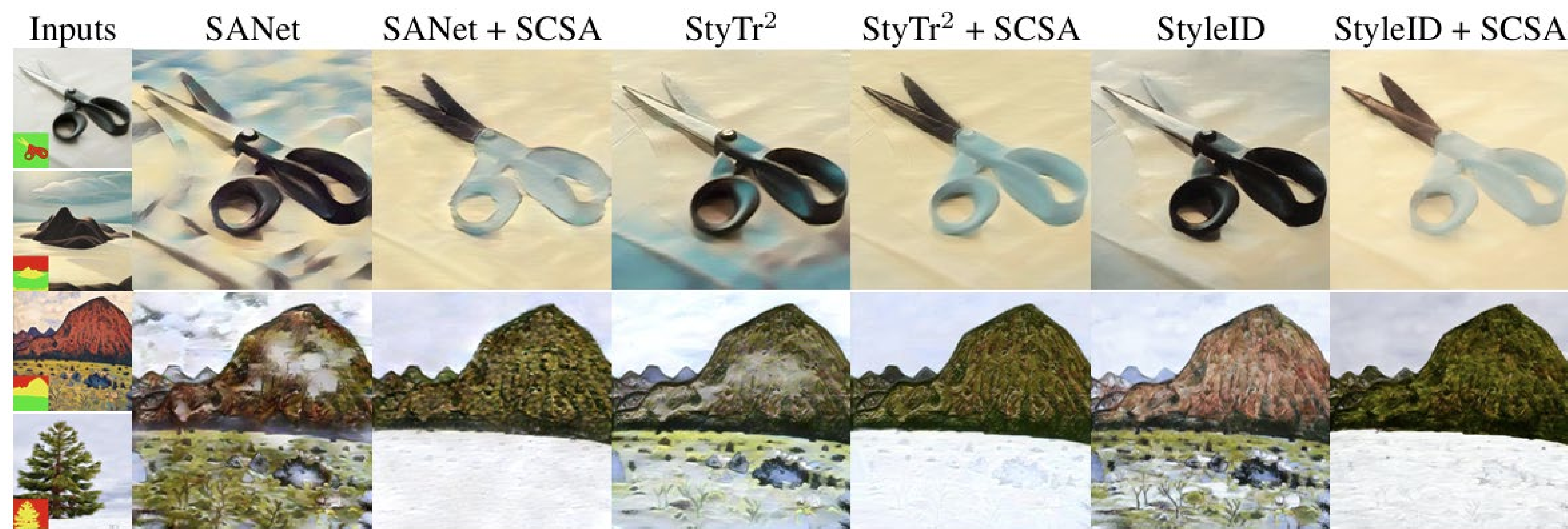
Experimental Results — Ablation Study

➤ Effect of SCA, SSA, and S-AdaIN

Content	Style	SANet + SCSA	- SSA	- SCA	- S-AdaIN	StyTR ² + SCSA	- SSA	- SCA	- S-AdaIN	StyleID + SCSA	- SSA	- SCA	- S-AdaIN
													
													
SSL↓		0.8762	0.8840	0.9096	0.8769	1.2228	1.4157	1.5714	1.2832	1.2447	1.3302	2.2981	1.7558
FID↓		13.0788	13.0814	12.0170	12.0275	12.3963	13.2059	14.1252	12.4782	12.4497	13.0826	15.4705	13.0359
CFSD↓		0.0874	0.0937	0.0994	0.0922	0.0705	0.0694	0.1668	0.0657	0.1178	0.1066	0.3402	0.0958

Experimental Results — Additional Analysis on Semantic Map

➤ Semantic style transfer based on semantic map



Contributions

- We reveal that attention-based arbitrary style transfer methods struggle with content and style images with the same semantics and identify the root causes for their subpar performance.
- We propose a semantic continuous-sparse attention, dubbed SCSA, that can extend attention-based arbitrary style transfer to arbitrary semantic style transfer in a plug-and-play manner.
- We conduct extensive experiments and comparisons to demonstrate the effectiveness and generalization of SCSA, enabling attention-based arbitrary style transfer methods (CNN-based, Transformer-based, and Diffusion-based approaches) to perform arbitrary semantic style transfer while preserving and even enhancing the original stylization.



Thank You for Listening!