3S-NET: ARBITRARY SEMANTIC-AWARE STYLE TRANSFER WITH CONTROLLABLE ROI CHOICE

Bingqing Guo, Pengwei Hao

Queen Mary University of London

Queen Mary
University of London

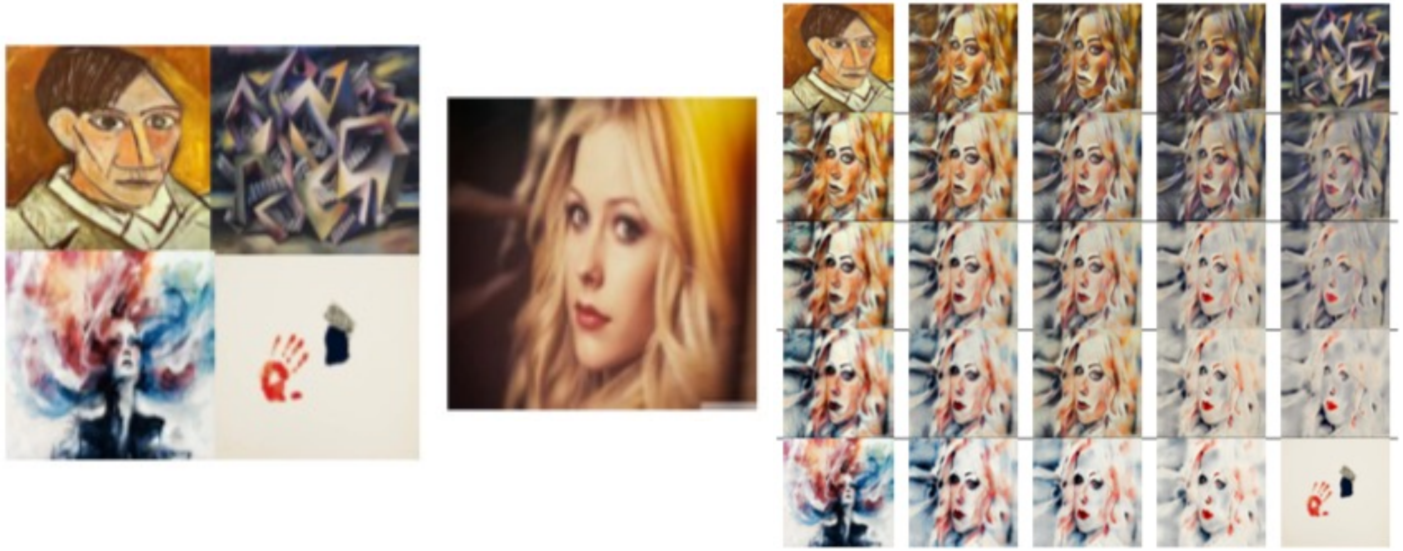
The Introduction of Style Transfer

Style transfer：**re-rendering a natural image by a selected artistic style.**

Style image | Content image | Style transfer results

Single style transfer

Multi-style transfer



Neural Style Transfer

- Optimization-based NST: generalize the transfer problem as a set of optimization problems [1];
- Training a feed-forward network: capture style patterns by proposed modules in the architecture [2, 3];
- Feature transformations-based methods: real-time and highly efficient; balance the efficiency and the visual quality of results [4].

[1] Gatys L A, Ecker A S, Bethge M. Image style transfer using convolutional neural networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2414-2423.
[2] Huang Z, Zhang J, Liao J. Style Mixer: Semantic-aware Multi-Style Transfer Network[C]//Computer Graphics Forum. 2019, 38(7): 469-480.
[3] Wu X, Hu Z, Sheng L, et al. StyleFormer: Real-Time Arbitrary Style Transfer via Parametric Style Composition[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 14618-14627.
[4] Li Y, Fang C, Yang J, et al. Universal Style Transfer via Feature Transforms[C]//NIPS. 2017.

Main research problem statement

- **Motivation:** Single style transfer is very popular recently. Most multi-style transfer techniques just linearly combine the style features with weights but overlook the semantic relations between the objects in the content image and style images.
- **Tasks:** an adaptive and flexible semantic-aware style transfer method that only needs transformation in feature space.
- **Key idea:** utilize different information encoded in each channel of features to perform effective semantic-aware style transfer (capable of single/multi/ROI style transfer).

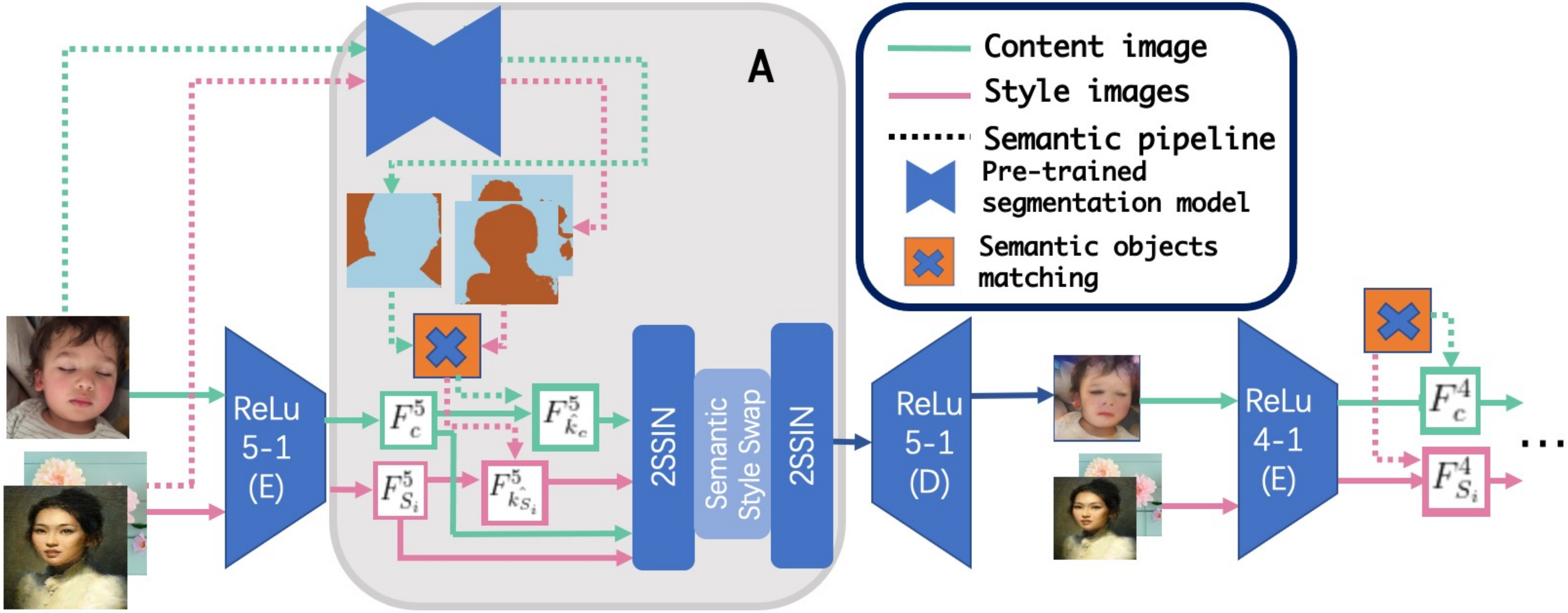
Key Contribution

- **1:** We design a framework called 3S-Net which is adaptive to meet multiple requirements. It also allows user-participation. Users can choose Region of Interest (ROI) to perform semantic-aware style transfer.

- **2:** We propose Two-Step Semantic Instance Normalization (2SSIN) and Semantic Style Swap. They utilize different information encoded in each channel of features to perform effective semantic-aware style transfer.

- To our best knowledge, our method is the first to achieve semantic-aware style transfer without training a model.

The Framework of Our Method



Methods Explanation

Semantic-driven Channel Descriptors：

The contribution $P^l_{\hat{k}}$ of each channel towards the group $\hat{k}$ is calculated as:

$$P^l_{\hat{k}} = \frac{1}{H^l W^l} \sum_{H^l, W^l} (F^l)^2 \odot u^l_{\hat{k}}, \qquad (1)$$

where $F^l \in \mathbb{R}^{1\times C^l \times H^l \times W^l}$, $u^l_{\hat{k}} = \Phi(u_{\hat{k}})$, the interpolation function is used for size matching. The standardization transform is added to F^l, so that F^l has zero mean and unit variance and $P^l_{\hat{k}} \in [0,1]^{1\times C^l}$.

With appropriate scale $\eta = \frac{1}{max(P^l_{\hat{k}})}$, we adjust the maximum of $P^l_{\hat{k}}$ to be 1. Then the weighted descriptors $Q^l_{\hat{k}}$ of $\hat{k}$ are obtained:

$$Q^l_{\hat{k}} = \eta P^l_{\hat{k}} \qquad (2)$$

Queen Mary University of London

Methods Explanation

Two-step Semantic Instance Normalization (2SSIN) ：

Step-1:

Channel-wise Semantic AdaIN:

Let $\hat{k_c}$ and $\hat{k_{S_i}}$ be two groups which correspond to each other, $u^l_{\hat{k_c}}$ and $u^l_{\hat{k_{S_i}}}$ indicate the masks of the l-th layer towards the content and the style groups respectively. $Q^l_{\hat{k_c}}$ and $Q^l_{\hat{k_{S_i}}}$ are the semantic channel weights of c and S_i, the process of channel-wise semantic AdaIN is shown as below:

$$f^l_{\hat{k_c},\hat{k_{S_i}}} = \gamma^l_{S_i} \cdot Q^l_{\hat{k_{S_i}}} \frac{F^l_{\hat{k_c}} - \beta^l_c \cdot Q^l_{\hat{k_c}}}{\gamma^l_c \cdot Q^l_{\hat{k_c}}} + \beta^l_{S_i} \cdot Q^l_{\hat{k_{S_i}}}, \tag{3}$$

$$F^l_{\hat{k_c}} = F^l_c \odot u^l_{\hat{k_c}}, \tag{4}$$

where $\beta, \gamma \in \mathbb{R}^{1\times C^l}$ represent the means and the standard deviations of the channels.

Methods Explanation

Two-step Semantic Instance Normalization (2SSIN) ：

Step-2:

Group-wise Semantic AdaIN:

The mean $\beta^l_{\hat{k}}$ and standard deviation $\gamma^l_{\hat{k}}$ of group $\hat{k}$ are calculated as below:

$$\beta^l_{\hat{k}} = \sum_{C_i \in \hat{k}} \frac{Q^l_{\hat{k},C_i} \cdot \beta^l_{C_i}}{\sum_{C_i \in \hat{k}} Q^l_{\hat{k},C_i}}, \gamma^l_{\hat{k}} = \sum_{C_i \in \hat{k}} \frac{Q^l_{\hat{k},C_i} \cdot \gamma^l_{C_i}}{\sum_{C_i \in \hat{k}} Q^l_{\hat{k},C_i}}, \tag{5}$$

$$F^l_{\hat{k_c},\hat{k_{S_i}}} = \gamma^l_{\hat{k_{S_i}}} \cdot \frac{f^l_{\hat{k_c},\hat{k_{S_i}}} - \beta^l_{\hat{k_c}}}{\gamma^l_{\hat{k_c}}} + \beta^l_{\hat{k_{S_i}}}, \tag{6}$$

where $F^l_{\hat{k_c},\hat{k_{S_i}}}$ is the final transferred feature and only has value in the region of the semantic group, $\hat{k_c}$,

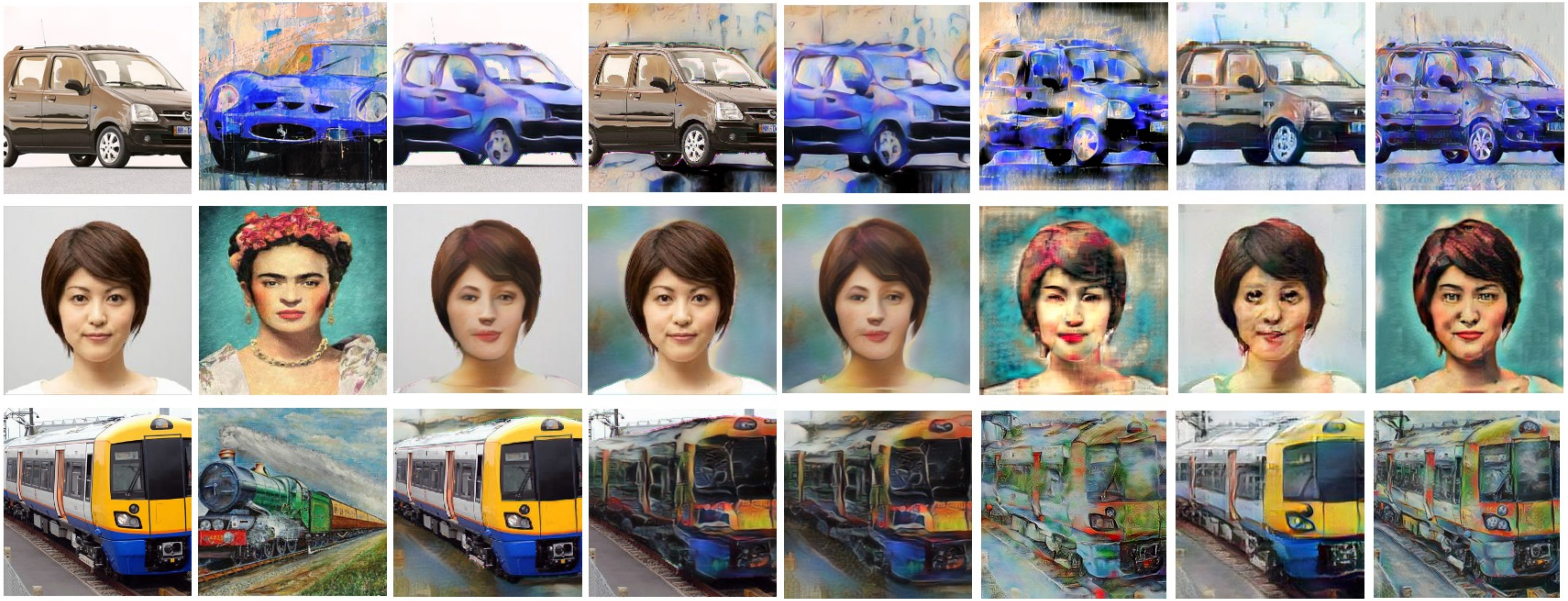
$$F^l_{cs} = \sum F^l_{\hat{k_c},\hat{k_{S_i}}} + \bar{F}^l_c, \tag{7}$$

Methods Explanation

Semantic Style Swap:

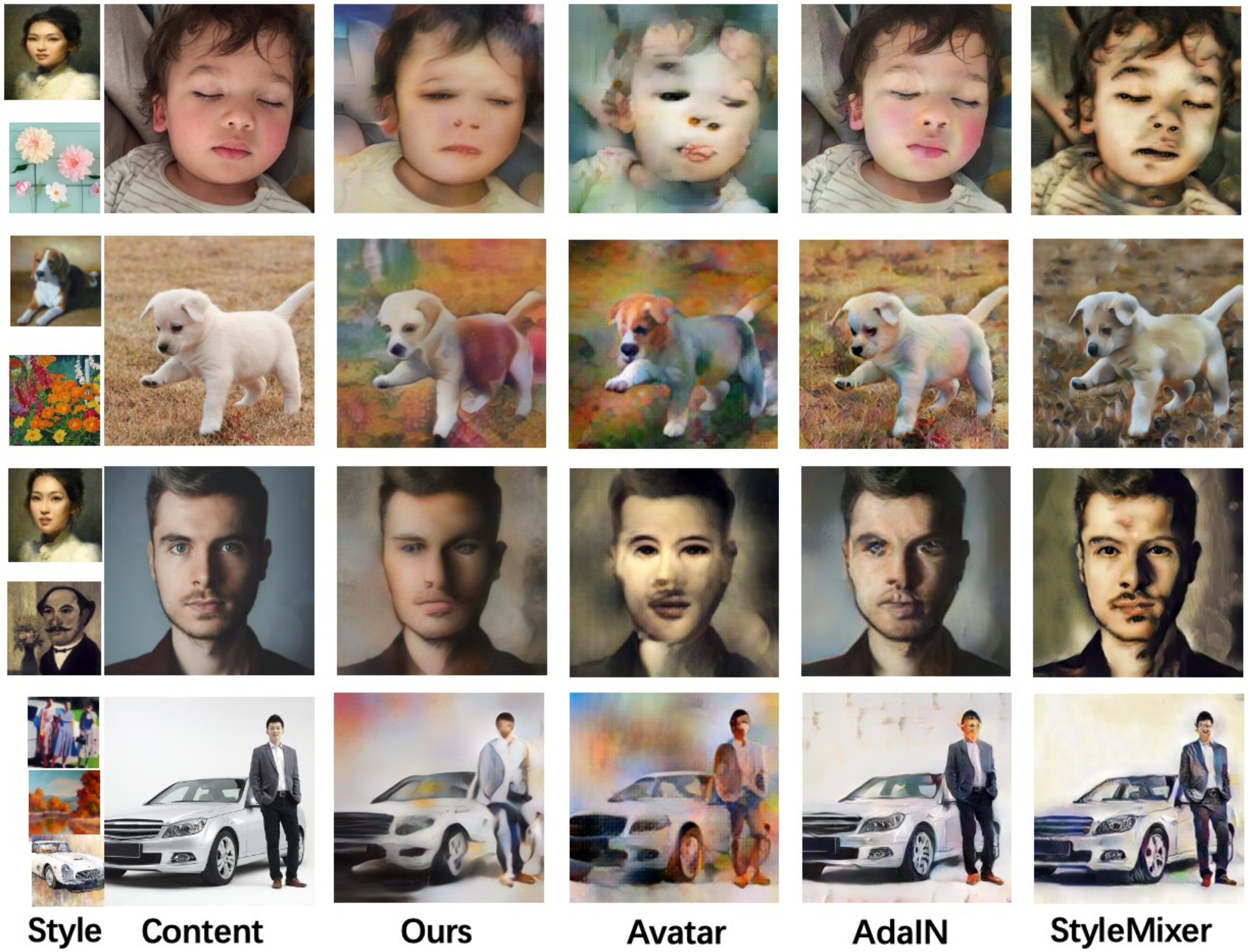
The improved Semantic Style Swap utilizes $u_c^{\hat{k}_c}$ and $u_{S_i}^{\hat{k}_{S_i}}$. $u_{S_i}^{\hat{k}_{S_i}}$ helps to guarantee that generating the patches from the regions of $\hat{k}_{S_i}$. So that when transferring the patches from the style image to content image, only patches of the specific $\hat{k}_{S_i}$ are transferred to the region of $\hat{k}_c$, which strengthens the semantic texture of the stylized features.

Qualitative Evaluation: Sigle Style Transfer



Content Style Ours ROI (A) Ours ROI (B) Ours (A+B) WCT AdaIN StyleMixer

Qualitative Evaluation: Multi-Style Transfer



More Examples