Research on Image Style Transfer Method Based on Semantic Adaptive

MA Chi, Wang Shaofan, Hu Hui

Abstract—**This paper proposes a novel image style transfer technique based on semantic adaptation to address the problems of content image representation and semantic information loss during the image style transfer process. Specifically, Two modules make up the method: the representation transfer module and the semantic transfer module. The representation transfer module extracts the content image's representation features through context coding. The semantic transfer module extracts the content image's semantic features by generating an adversarial network. These modules effectively preserve the content image's information at both the representation and semantic levels. In comparative experiments with various image style transfer methods, our proposed method achieves significantly better results. Thus, the proposed method effectively retains the representation and semantic information of content images during the style transfer process.**

Index Terms—**Deep learning, image style transfer, generative adversarial networks, context encoding, semantic adaptation**

I. Introduction

IMAGE style transfer has remained a focal point of sustained scholarly attention, with related studies known as Non-photorealistic rendering (NPR), which is used for the stylization of two-dimensional images[1]. The process of transforming a content image into one with a specific artistic style is, in essence, image style transfer. The stylized image integrates the content image's semantic cues with the style image's textural and chromatic details. Although this type of technology can imitate certain specific artistic styles and produce well-styled images, the models constructed are usually too complex. As a result, they cannot be flexibly applied to various arbitrary styles and cannot semantically model complex textures. With the rapid advancement of deep learning technologies in recent years [23, 24], numerous methods for image style transfer have emerged and demonstrated remarkable achievements.

Currently, image style transfer can be divided into many different types. Among these, the more popular methods include those based on the Gram matrix [2-5] and Generative Adversarial Networks (GAN) [6-10]. For the former, in 2016, The image optimization-based image style transfer method was first proposed by Gatys et al. [11]. This method separates and recombines content and style images through Convolutional Neural Networks (CNN) and their Gram matrices. To accelerate this method, Johnson et al. [12] and Ulyanov et al. [13] proposed a model-based image style transfer method. Chen et al. [14] explicitly separated the content and style information of the image using filters, achieving multiple style transfers of the image. By implementing a cross-training approach, Zhou et al. [15] achieved improved retention of content image details in stylized results. However, in an effort to preserve the content image's detailed information, these methods have no choice but to do so at the cost of reducing the degree of stylization. For the latter, Isola et al. [16] proposed a universal model for image-to-image transfer tasks, Pix2Pix. This model is based on conditional adversarial networks [19] and can effectively address a class of issues related to image transformation. Sun et al. [18] proposed a new generative adversarial network model by introducing and optimizing the squeeze excitation residual block and loss function. This model is capable of fully expressing the details and facial features of cartoon faces. Yue et al. [25] proposed a GANs-based method for filling blank bands in electric logging images. This method effectively improved the naturalness of the filled images and the continuity of geological structure textures by introducing dilated convolutional layers and adversarial training.

However, the neural network techniques currently used for image stylization have limitations. The method based on the Gram matrix tends to lose information of the content image. Moreover, the method based on Generative Adversarial Networks (GAN) requires paired datasets. Addressing these issues, this paper presents an image style transfer method based on semantic adaptation.

- We have analyzed the representation of image features in convolutional neural network layers at different depths. This allows us to determine the convolutional layers needed for extracting content and style features.
- The Representation Transfer Module is proposed for stylizing images. This module is a matrix-based method that introduces a context encoding module to handle complex textures and color information, addressing issues such as object edge distortion and color overlay. Additionally, instance normalization is added after the convolutional layers, enabling the network to converge more rapidly during training.
- The Semantic Transfer Module is proposed. This module is a method based on conditional Generative

Manuscript received Nov. 17, 2024; revised Sep. 09, 2025.

This paper is supported by the Foundation of Guangdong Educational Committee under the Grant No. 2022ZDZX4052 and 2021ZDJS082.

MA Chi is an associate professor of the School of Computer Science and Engineering, Huizhou University, Huizhou, 516007, China (e-mail: machi@hzu.edu.cn).

Wang Shaofan is an engineer of the Derkee Information Co., Ltd, Shenzhen, 518000, China (corresponding author. Tel.: +86-137-1488-5357; e-mail: wsf19961230@163.com).

Hu Hui is a lecturer of the School of Computer Science and Engineering, Huizhou University, Huizhou 516007, China (e-mail: Huhui@hzu.edu.cn).

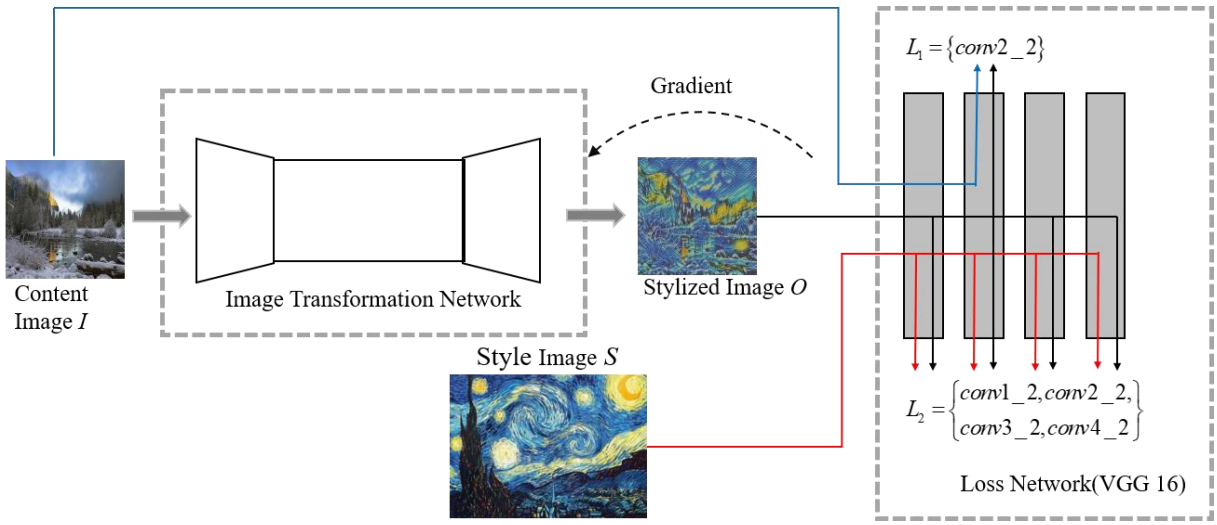


Fig. 1. Representation transfer module

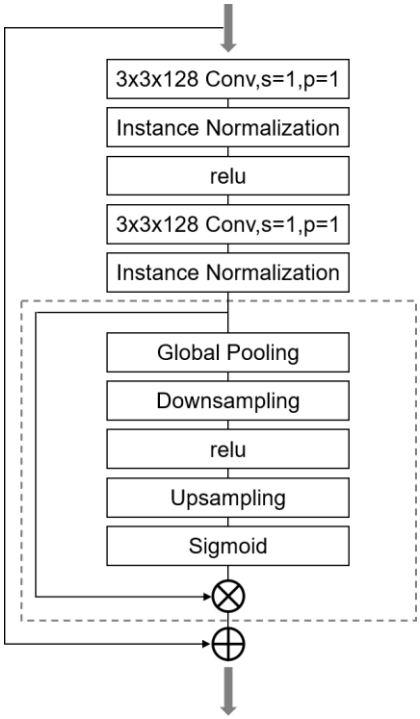


Fig. 2. Residual block structure and parameters in image conversion network

Adversarial Networks (GANs). o more effectively retain the the content image's semantic information , a semantic extractor is introduced. Through the adaptive optimization of the min-max value between the generator and the discriminator, the module learns the style image's textural and chromatic details. Meanwhile, the semantic extractor optimizes the generator by using the content image's high-level semantic features as a constraint condition.

- Finally, to boost the quality and authenticity of the stylized images, the Representation Transfer Module and the Semantic Transfer Module are integrated sequentially.

II. Representation Transfer Module

The Representation Transfer Module is composed of a loss network and an image transformation network. The image transformation network is constructed with a structure of downsampling, residual blocks, and upsampling; the loss network utilizes the VGG16 network's corresponding layers, as shown in Fig. 1. In the figure, the blue and red arrows represent the process of the content image and style image being input into the network, respectively. The black arrows represent the residuals between the stylized image and the other input images, respectively.

The module takes content image I as the input parameter and generates the output result stylized image O after internal processing. Subsequently, the content image I , style image S , and stylized image O are fed into the loss network, which makes the stylized image O form content and style losses with the corresponding layers of the loss network (VGG16) and their corresponding Gram matrices. The image transformation network is updated through gradient descent optimization, the stylized image O is equipped with content and style representations.

To generate more delicate stylized images, cross-channel and within-channel context encoding modules were introduced into the residual blocks and upsampling parts of the image transformation network. These modules introduces extensive contextual information into other local convolutional filters.

The introduction of the inter-channel context encoding module is an improvement to the residual blocks in the image transformation network. As shown in Fig. 2, the structure within the dashed box represents the inter-channel context encoding module. This module enables the network to adaptively predict the relative importance of feature map channels. Consequently, the image transformation network is better able to extract the content image's semantic cues while preserving the style image's textural and chromatic details.

The introduction of the intra-channel context encoding module is an improvement in the upsampling process in the image transformation network. This module employs a pyramid module to perform pooling operations at different scales on each feature map, allowing for the integration of multi-scale spatial information within the channels of the feature maps. This enables the image transformation network

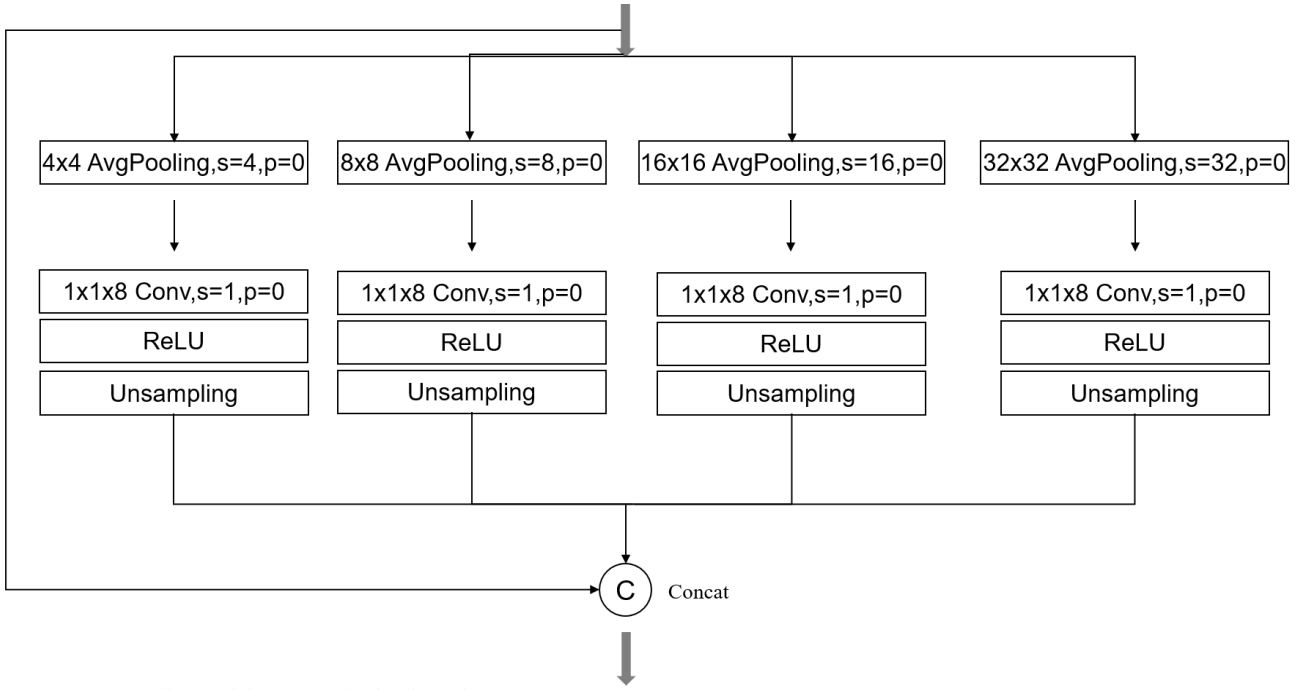


Fig. 3. Context encoding module structure in the channel

to generate stylized images that retain more details from the content image, as depicted in Fig. 3.

For image style transfer, each image should be considered as a different domain, and in order to maintain the independence between different image instances, an instance normalization[21] (IN) follows every convolutional layer. This operation promotes the improvement of network convergence efficiency during the model training process, thereby enhancing the capability of the image style transfer method.

The loss function of the Representation Transfer Module is composed of three parts: content, style and total variation loss. The content loss can be used to measure the difference between the content image and the stylized image at the feature map level, as shown in Eq. (1).

$$\mathcal{L}_{content}(O,I)=\frac{1}{H_l W_l C_l}\sum_{l\in\{l_c\}}\| F^l(I)-F^l(O)\|^2 \quad (1)$$

Among them, $\{l_c\}$ represents conv2_2 layer of the VGG16 network, whose function is to represent content information. H_l, W_l, and C_l denote the height, width, and number of channels, respectively, of the feature map obtained after the image is processed by the l-th convolutional layer of the VGG16 network. $F^l(I)$ and $F^l(O)$ represent the feature maps corresponding to the content image and the stylized image after passing through the l-th convolutional layer of the VGG16 network.

The style loss function represents the stylistic representation between the style image S and the stylized image O, as shown in Eq. (2).

$$\mathcal{L}_{style}(O,S)=\frac{1}{H_l W_l C_l}\sum_{l\in\{l_s\}}\| G(F^l(S))-G(F^l(O))\|^2 \quad (2)$$

In this context, $\{l_s\}$ represent the conv1_2, conv2_2, conv3_2, and conv4_2 layers of the VGG16 network.

$G(F^l(S))$ and $G(F^l(O))$ denote the Gram matrices of the feature maps extracted from the style image S and the stylized image O, respectively, after they are forwarded through the l-th convolutional layer, as shown in Eq. (3).

$$G^l(F)=F^l(F^l)^T, F^l\in R^{C_l\times HW} \quad (3)$$

In this context, F^l denotes the feature map produced by the l-th convolutional layer, reshaped from $N_l\times H\times W$ to $N_l\times HW$, where N_l, H, and W denote the number of channels, height, and width, respectively. $G^l(F)$ is a global statistic of the feature map, serving as a representation of the image's style.

For the purpose of make the generated stylized images smoother, this paper introduces Total Variation (TV) loss, which regularizes high-frequency components by penalizing large one-pixel differences in both spatial directions, as defined in Eq. (4).

$$\mathcal{L}_{TV}(x)=\sum_{i,j}\| x_{i,j+1}-x_{i,j}\|^2+\| x_{i+1,j}-x_{i,j}\|^2 \quad (4)$$

In which $x_{i,j}$ represents the value of the corresponding pixel point in the stylized image O. Essentially, high-frequency components are edge detectors, as shown in Fig. 5.

The representation transfer module's total loss is formed as the weighted sum of the content loss, style loss, and total variation loss, as shown in Eq. (5).

$$\mathcal{L}_{total}=\alpha\mathcal{L}_{content}+\beta\mathcal{L}_{style}+\lambda\mathcal{L}_{TV} \quad (5)$$

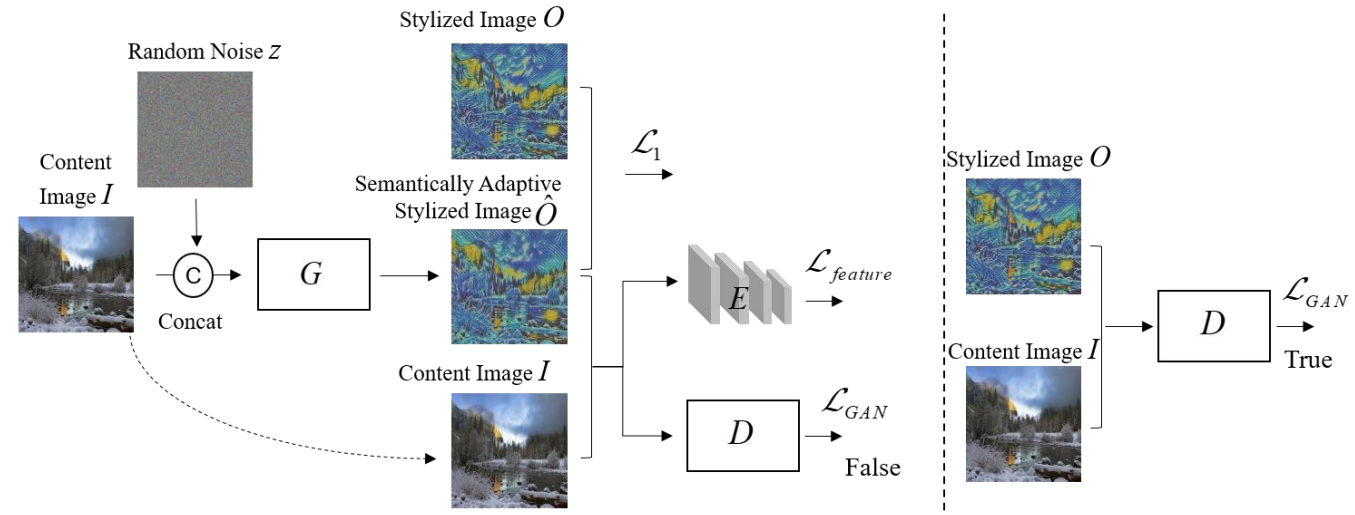


Fig. 4. Semantic transfer module

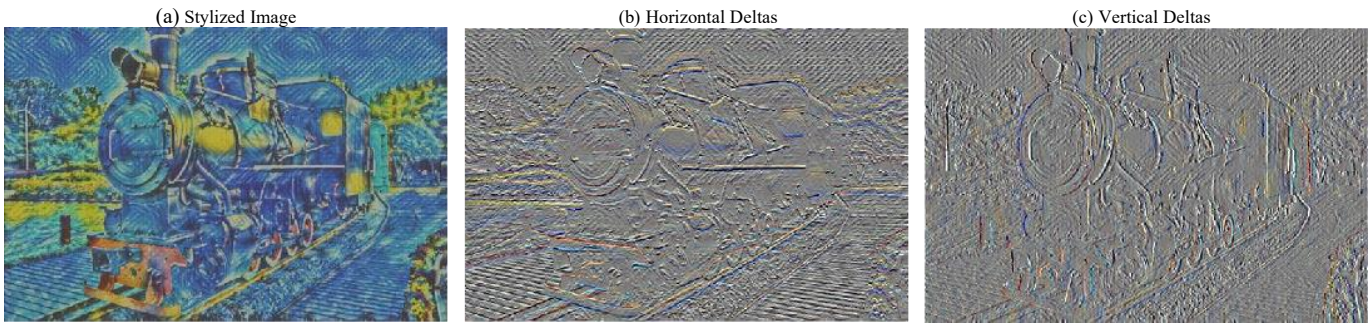


Fig. 5. Stylized image and its horizontal and vertical components

In which α , β and λ are weight coefficients, respectively 1, 1×10^5, and 1×10^{-6}.

III. Semantic Transfer Module

The representation transfer module in the image transformation network introduces a context encoding module that facilitates the network's learning of more content image details. However, this process merely constrains the stylized image's content and style information, without using the semantic representation of the content image for stylization, resulting in the stylized image lacking semantic representation. To solve this problem, based on the aforementioned method, further processing is performed on the stylized image to construct a semantic transfer module, as shown in Fig. 4. This module is a conditional generative adversarial network consisting of a generator, a discriminator, and a semantic extractor.

The generator G adopts the "U-Net" network structure [20], taking the content image I and random noise $z \sim N(0,1)$ as inputs, and outputs the semantically adaptive stylized image $\hat{O} = G(z, I)$. The discriminator D adopts the "PatchGAN" network structure; it takes the semantically adaptive stylized image $\hat{O}$ or the stylized image O as input, conditioned on the content image I , and outputs a probability value in the range [0, 1]. The semantic extractor E computes a feature loss using the conv4_2 layer of the VGG16 network, it takes the semantically adaptive stylized image $\hat{O}$ and the content image I ,and outputs the distance between their respective conv4_2 feature maps.

The Generative Adversarial Network (GAN) model, as an unsupervised learning method, fundamentally the competitive mechanism is established between the generator G and the discriminator D . The adversarial loss $\mathcal{L}_{GAN}$ is shown in Eq. (6).

$$\begin{aligned}\mathcal{L}_{GAN} &= \mathbb{E}_{O\sim P_{data}(O),I\sim P_{data}(I)}[\log D(O,I)] \\ &+ \mathbb{E}_{z\sim N(0,1),I\sim P_{data}(I)}[\log(1-D(G(z,I),I))]\end{aligned} \tag{6}$$

In which, $O \sim P_{data}(O)$ denotes sampling from the stylized image O ; $I \sim P_{data}(I)$ denotes sampling from the content image I ; $z \sim N(0,1)$ denotes obtaining samples from a normal distribution that follows a mean of 0 and a variance of 1.

For the purpose of enhance the semantic representation of the semantically adaptive stylized image $\hat{O}$ for the content image I , by minimizing the feature loss $\mathcal{L}_{feature}$ between the content image I and the semantically adaptive stylized image $\hat{O}$, it enables the generator to learn the content image's semantic cues, as shown in Eq. (7).

$$\mathcal{L}_{feature} = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\| E(I_{ij}) - E(\hat{O}_{ij}) \|^2 \tag{7}$$

In which, m and n denote the width and height of the

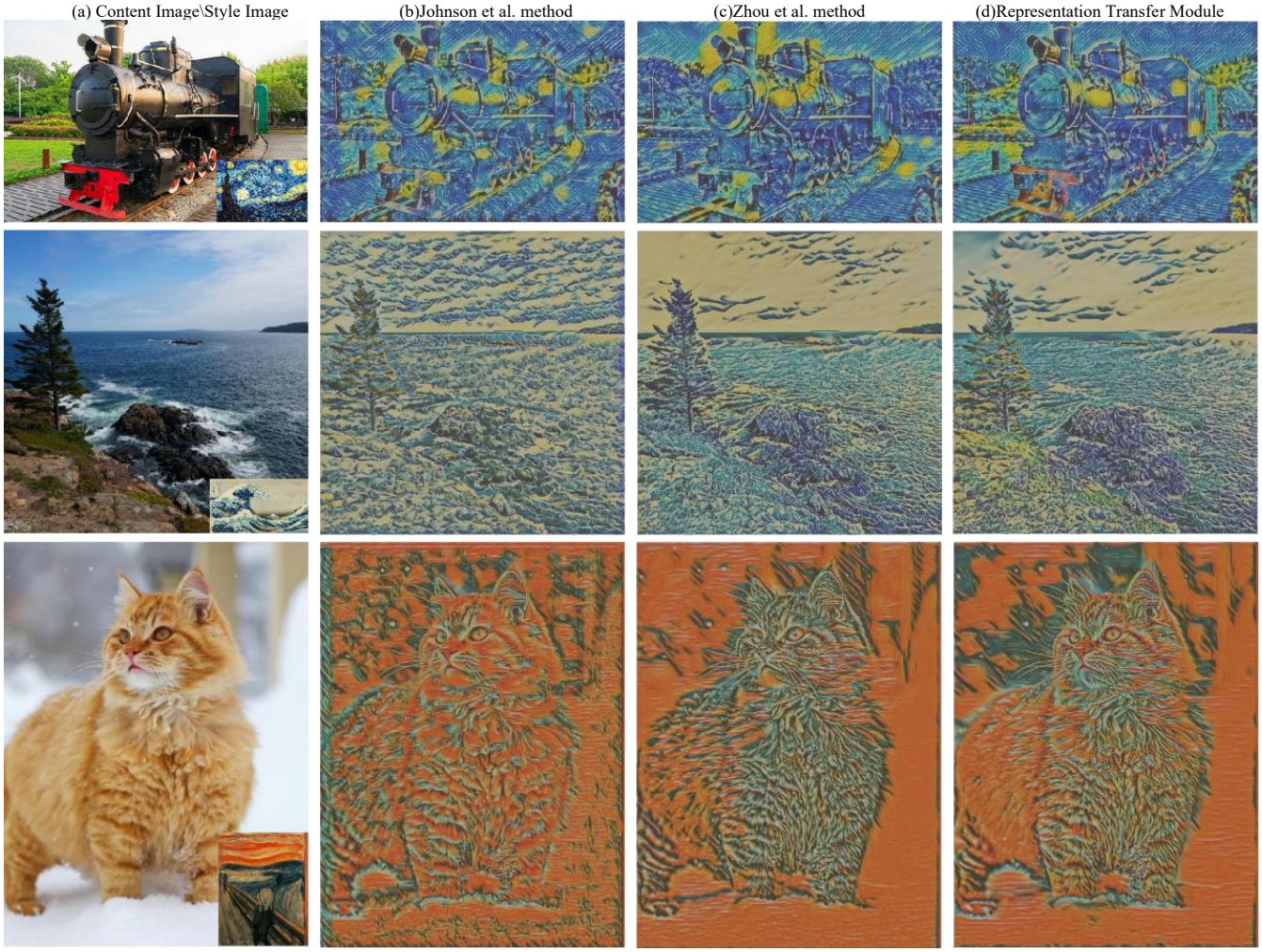


Fig. 6. Comparison of stylized images of different algorithms

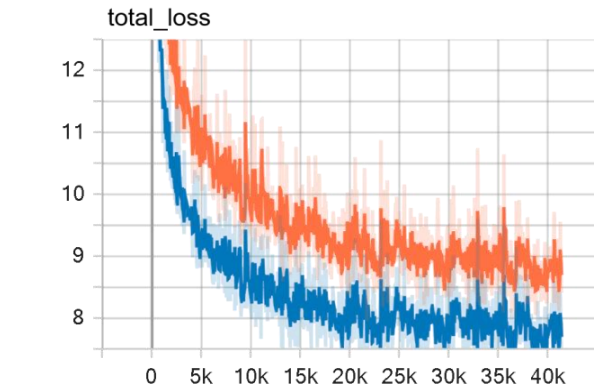


Fig. 7. Changes in the total loss of different algorithms

feature map, respectively.

$\hat{O}_{ij}$ denotes the pixel value of the semantically adaptive stylized image $\hat{O}$ at (i, j), and I_{ij} denotes the pixel value of the content image I at (i, j).

By minimizing the O loss for the stylized image $\hat{O}$ and the semantically adaptive stylized image $\mathcal{L}_1$, the generator G produces clearer images, as shown in Eq. (8).

$$\mathcal{L}_1 = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}|O_{ij} - \hat{O}_{ij}| \tag{8}$$

The total loss of the semantic transfer module is the weighted sum of the generative adversarial loss $\mathcal{L}_{GAN}$, the feature loss $\mathcal{L}_{feature}$, and the $\mathcal{L}_1$ loss, as shown in Eq. (9).

$$\mathcal{L}_{total} = \gamma\mathcal{L}_{GAN} + \delta\mathcal{L}_{feature} + \mu\mathcal{L}_1 \tag{9}$$

In which γ, δ, and μ are weight coefficients, respectively 1, 10, and 100.

IV. Experiment

For the purpose of verify the superiority of the representation transfer module in image style transfer, comparative experiments were conducted with methods described in references [12] and [15]. Subsequently, an ablation study was conducted on the representation transfer module to further confirm the efficacy of the context encoding module. Finally, taking the representation transfer module as a foundation and integrating the semantic transfer module, comparative experiments were performed with the representation transfer module and reference [15], to illustrate the proposed method's excellence in image style transfer tasks.

A. Implementation Details

The implementation of the method proposed in this paper

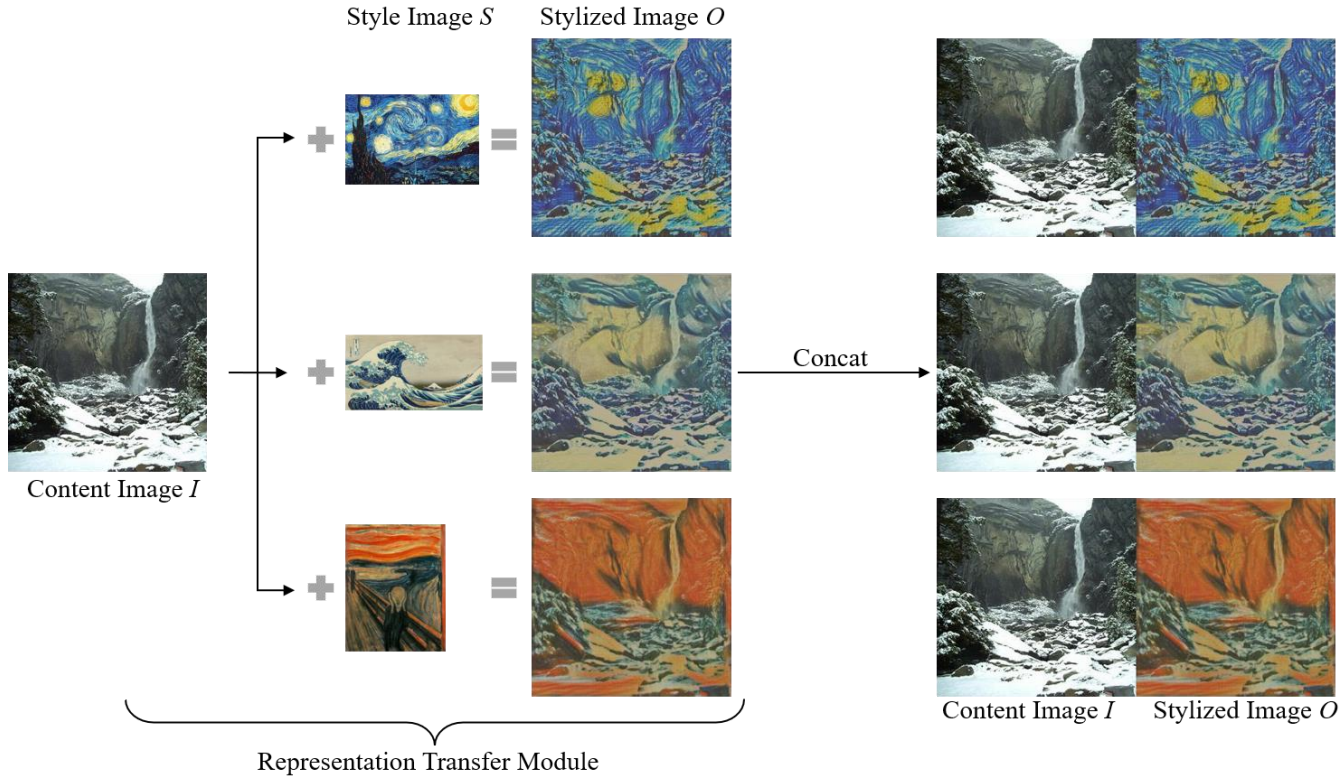


Fig. 8. Data set construction process

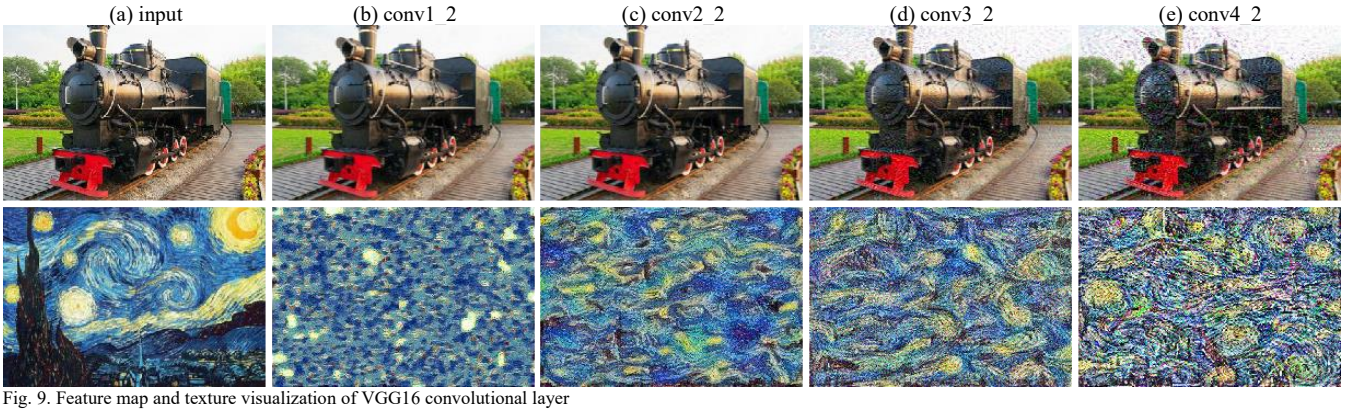


Fig. 9. Feature map and texture visualization of VGG16 convolutional layer

relies on TensorFlow version 2.4.1, with an NVIDIA RTX 3070 graphics card with 8GB of memory, and a batch size of 4 for training the network model. The representation transfer module and the semantic transfer module are optimized with the Stochastic Gradient Descent (SGD) algorithm. The Adam optimizer is employed, with exponential decay rates for moment estimation set to 0.9 and 0.99, respectively. The initial learning rate for updating the image transformation network in the representation transfer module is set to 1×10^{-3}, and the initial learning rates for updating the generator and discriminator in the semantic transfer module are set to 2×10^{-4}.

B. Representation Transfer Module

1) Loss Variation and Analysis of Different Algorithms: To analyze the impact of incorporating context encoding modules and instance normalization into the image transformation network for image style transfer, the Microsoft COCO dataset [22], a widely used and authoritative dataset in computer vision, is used as the training dataset for content images. This dataset contains a diverse set of 80,000 images. For style images, the WikiArt dataset, which has a rich artistic style, is used, and the training time is approximately 2 hours.

This paper employs the Tensorboard X tool to visualize the changes in total loss, as shown in Fig. 7. The red and blue lines represent the changes in total loss of Johnson et al. [12] and the representation transfer module, respectively. As the number of iterations increases, the loss values decrease and the curves become smoother, completing the image style transfer process. From the different loss change curves in the figure, it is evident that incorporating context encoding modules and instance normalization operations into the image transformation network allows the model to converge faster and enhances the feature learning capabilities of the image transformation network.

2) Comparative Display of Stylized Images Using Various Methods: To verify the superiority of the representation transfer module, a comparative experiment was conducted to evaluate the image generation effects between the representation transfer module and the methods of Johnson et al. [12] and Zhou et al. [15]. The same content and style images were used, with each row corresponding to style images of starry, wave, and scream, respectively. The stylized images generated by different methods are shown

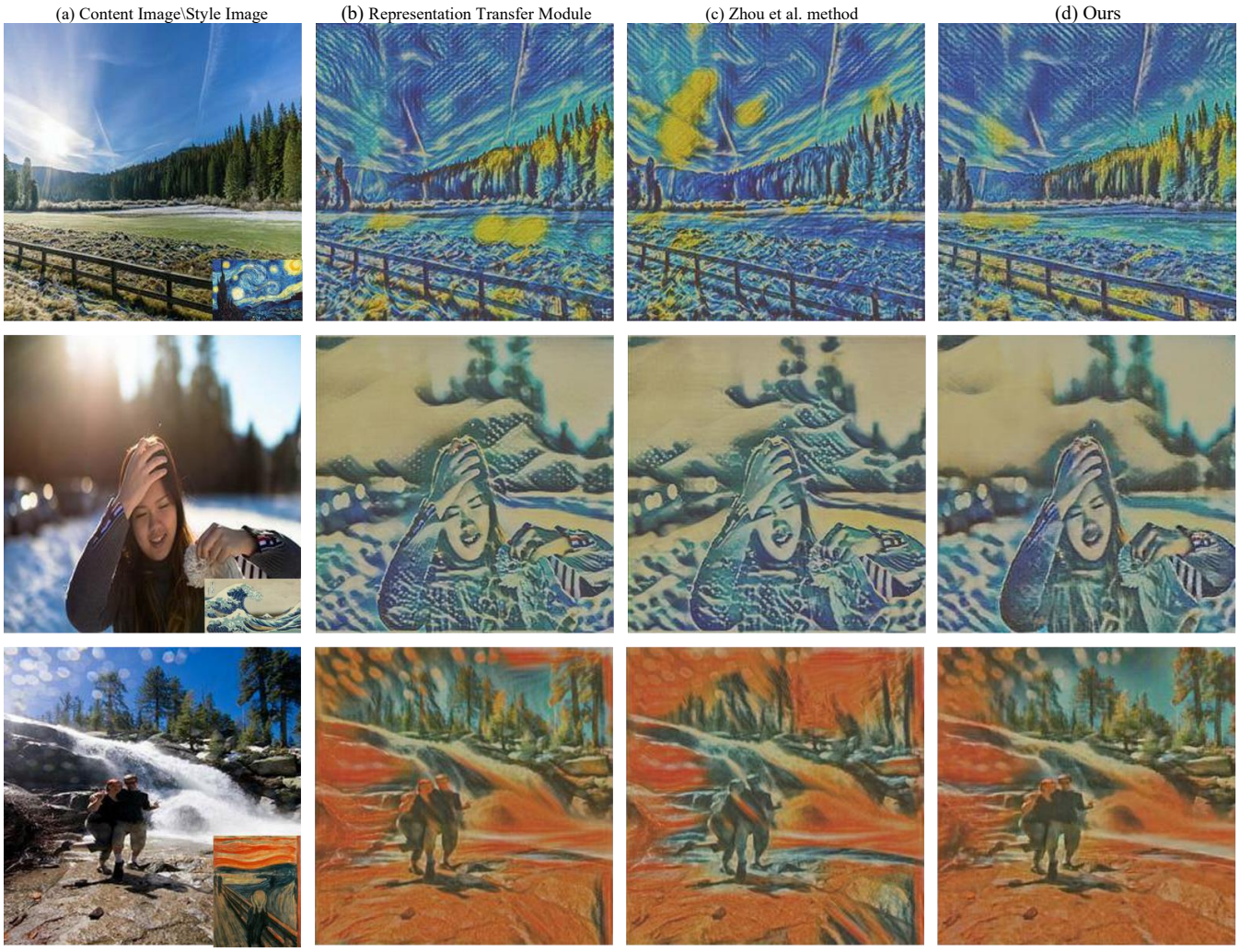


Fig. 10. Semantic adaptive optimization results display

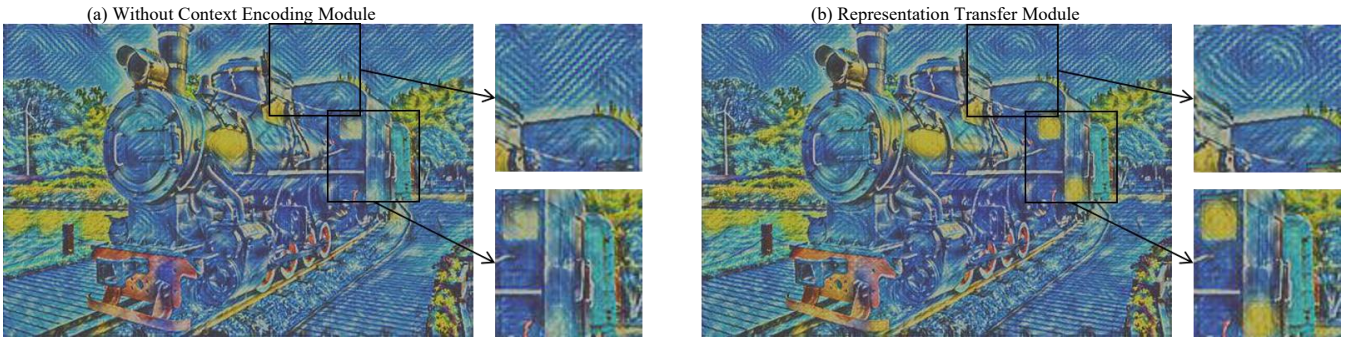


Fig. 11. Comparison of the details of the stylized image between the variant method and the method in this article

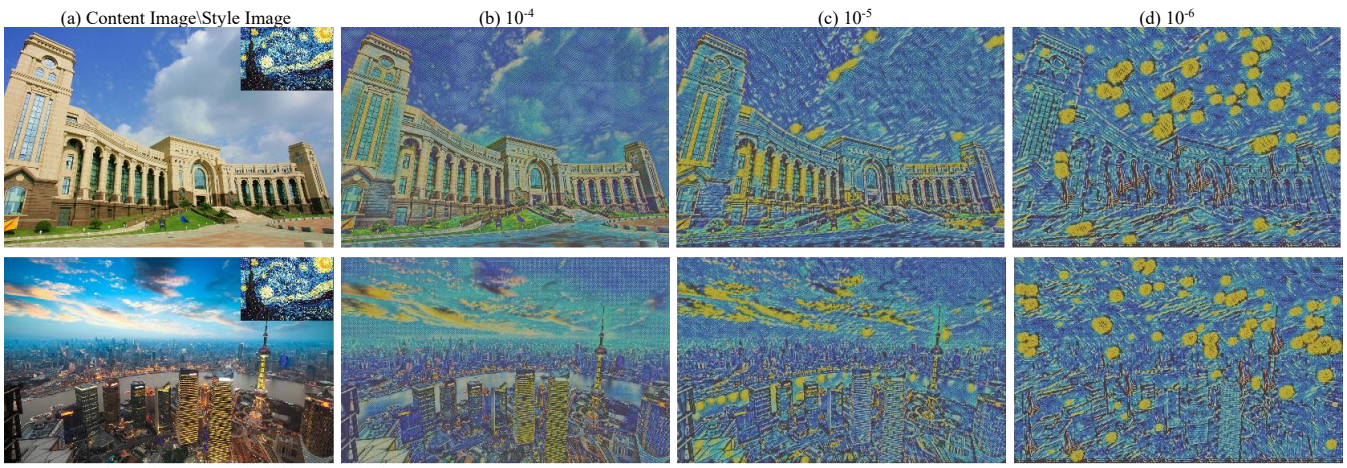


Fig. 12. Stylized images with different weight ratios

in Fig. 6.

As shown in Fig. 6, compared with Johnson et al.'s method, the stylized images generated by the representation transfer of objects and exhibit a more three-dimensional feel.Although the results produced by Zhou et al.'s method have a certain degree of three-dimensionality, their stylized images lack rich colors and miss details of the content images. Therefore, the introduction of context encoding modules and instance normalization operations can enable the network to better extract multi-scale spatial features from feature channels. In summary, the representation transfer module can preserve the details of content images more effectively and has a three-dimensional feel, thus generating better image stylization effects.

3) Analyze the feature extraction characteristics of different layers of the network: To explore the feature extraction characteristics of each layer in the VGG16 network pre-trained on the ImageNet dataset, this paper employs the backpropagation algorithm for feature visualization of each layer in the neural network. Specifically, high-level layers capture high-level semantic features, which can be employed to reconstitute the semantic content of the image. In contrast, low-level layers capture pixel-level details, allowing for the simple reconstruction of the original image's pixel values, as shown in Fig. 9.

4) Validating the Effectiveness of the Context Encoding Module: To additionally validate the functionality of the context encoding module, variant experiments were conducted on the representation transfer module. These experiments compared the representation transfer module with methods that do not include the context encoding module. A detailed comparison of the stylized images generated by both methods is shown in Fig. 11.

As shown in Fig. 11, the method without the context encoding module fails to fully render the stylized images during the image style transfer process. Specifically, the enlarged images reveal that the texture of the sky and the color rendering of the train windows in Fig. 11(b) are more pronounced. Therefore, while preserving the details of the content image, the context encoding module can also effectively integrate style features into the image.

5) Analyzing the impact of weights on the stylized images: To investigate the impact of the proportion of content and style losses on the stylized images, we adjusted the experiments by using different ratios of content loss to style loss α / β when calculating the total loss. The effects of the generated stylized images under different proportions of the two loss values are shown in Fig. 12. As depicted in the figure, we can achieve a stronger style effect in the generated images by reducing the loss weight ratio (increasing β).

C. Semantic Transfer Module

The representation transfer module uses a global correlation matrix to achieve image style transfer without needing paired datasets. However, the semantic transfer module, which is based on conditional generative adversarial networks (CGANs), requires paired datasets. Specifically, the summer2winter dataset [17] is employed as the content image, and the corresponding stylized images O are generated through the representation transfer module; then, the corresponding content images I and stylized images O are concatenated to construct the dataset, as shown in Fig. 8.

There are three datasets corresponding to the styles of starry, scream, and wave. Each dataset includes a training set with 2,000 images as well as a test set with 100 images.

1) Analysis of the Variations in Each Loss: The semantic transfer module, which is based on the conditional generative adversarial network framework and incorporates the representation transfer module, constitutes a semantic adaptive image style transfer method. It is trained through the adaptive optimization of the generator and the discriminator, together with the constraints imposed by the feature loss and the $\mathcal{L}_1$ loss. To analyze the impact of the semantic extractor on other loss variations, the Tensorboard X tool is used to visualize the $\mathcal{L}_{GAN}$, $\mathcal{L}_{feature}$, and $\mathcal{L}_1$ losses, respectively, and the training takes approximately 40 minutes. It should be noted that $\mathcal{L}_{GAN}$ is composed of the discriminator loss and the generator loss, corresponding to the loss variation processes in the figures 10(a) and (b). The value of log(2)=0.69 serves as a good reference point, as it indicates a perplexity of 2. On average, the discriminator has the same uncertainty for these two options. As shown in Fig. 14(a), the discriminator loss had been continuously decreasing before 10k, and after that, the loss value has been around 0.69, reaching an adaptive balance. This indicates that the semantically adaptive stylized image $\hat{O}$ and the stylized image O follow the same distribution, meaning they share the same texture style.

As shown in Fig. 14(c), the feature loss initially drops sharply and then rises gradually as the training progresses. This variation in loss describes the semantic differences between the content image I and the semantically adaptive stylized image $\hat{O}$. The sharp decrease indicates that when the semantically adaptive stylized image $\hat{O}$ has a weak style, its semantic information is close to that of the content image I . As training progresses, the gradual increase indicates that the semantically adaptive stylized image $\hat{O}$ loses some of the semantic information of the content image I . This further illustrates that the process of stylizing the content image I is also a process of semantic information loss. Fig. 14(c) indicates that the introduction of feature loss has partially preserved the semantic information of the content image I . Fig. 14(d) illustrates that as training progresses, the $\mathcal{L}_1$ loss continuously decreases, and the difference between the semantically adaptive stylized image $\hat{O}$ and the stylized image O becomes increasingly close, ultimately resulting in clearer generated images.

2) Comparison of Stylized Image Details: As shown in Fig. 13, to illustrate that the stylized images generated by the method integrating the semantic transfer module have greater advantages in the semantic representation of content images, a comparison is conducted between the results of the image style transfer method based on semantic adaptation (our method), the representation transfer module, and the method proposed by Zhou et al. [15]. The styles, from top to bottom, correspond to starry, wave, and scream. The figure clearly shows that our method provides a stronger semantic

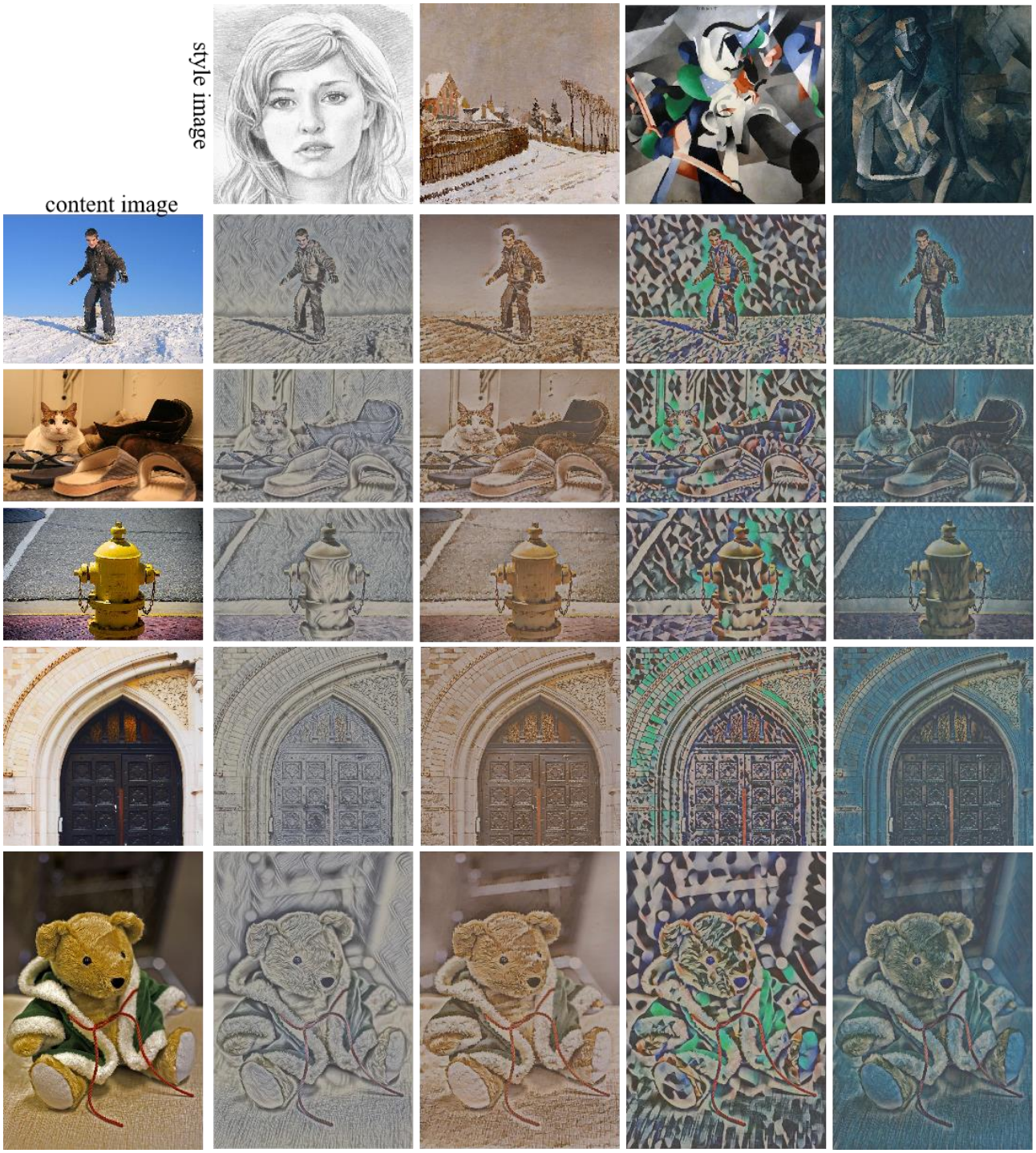
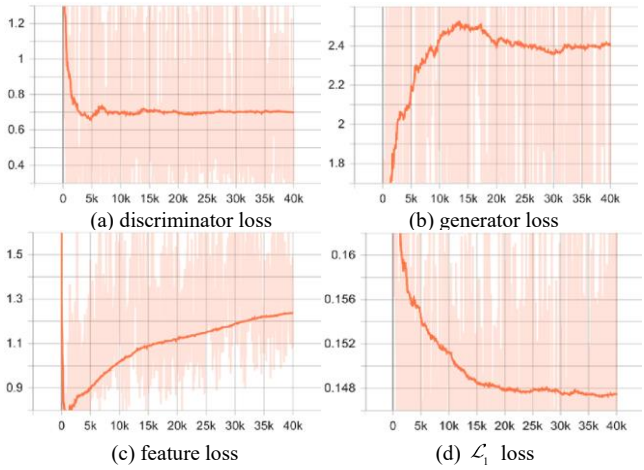


Fig. 13. Display of Image Generation Results Using the Ours Method



(a) discriminator loss (b) generator loss

(c) feature loss (d) $\mathcal{L}_1$ loss

Fig. 14. The changes of various loss values during the optimization process.

representation of the content image, highlighting its superiority.

V. Conclusion

This paper introduces a semantically adaptive image style transfer method that adaptively migrates the style of an image based on semantics. Specifically, the representation transfer module uses the pre-trained VGG16 network layers and their corresponding Gram matrices to constrain the image transformation network to produce stylized images. The context encoding module facilitates the image transformation network in better preserving and highlighting the fine-grained features of the content image. Subsequently, on the basis of the representation transfer module, a semantic

transfer module based on the framework of conditional generative adversarial networks is added, using the conv4_2 layer of the VGG16 network as a semantic extractor to extract semantic information, thus forming a semantically adaptive image style transfer method. In contrast to other methods, our method not only demonstrates superior performance in image style transfer but also addresses issues such as semantic loss, object edge distortion, and color overlay during the image style transfer process.

REFERENCES

[1] Jan Eric Kyprianidis, John Collomosse, Tinghuai Wang, and Tobias Isenberg, "State of the "Art": A Taxonomy of Artistic Stylization Techniques for Images and Video," IEEE Transactions on Visualization and Computer Graphics, vol. 19, no. 5, pp866-885, 2012.

[2] Falong Shen, Shuicheng Yan, Gang Zeng, "Neural Style Transfer via Meta Networks," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp8061-8069, 2018.

[3] Hui-Huang Zhao , Paul L. Rosin, Yu-Kun Lai, Mu-Gang Lin, Qin-Yun Liu, "Image Neural Style Transfer With Global and Local Optimization Fusion," IEEE Access, vol. 7, pp85573-85580, 2019.

[4] Wenjing Wang, Jizheng Xu, Li Zhang, Yue Wang, Jiaying Liu, "Consistent Video Style Transfer via Compound Regularization," Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, no. 7, pp12233-12240, 2020.

[5] Ali Uzpak, Abdelaziz Djelouah, Simone Schaub-Meyer, "Style Transfer for Keypoint Matching Under Adverse Conditions," 2020 International Conference on 3D Vision (3DV), pp1089-1097, 2020.

[6] Antonia Creswell, Tom White, Vincent Dumoulin, Kai Arulkumaran, Biswa Sengupta, Anil A. Bharath, "Generative Adversarial Networks: An Overview," IEEE Signal Processing Magazine, vol. 35, no. 1, pp53-65, 2018.

[7] Huiwen Chang, Jingwan Lu, Fisher Yu, Adam Finkelstein, "PairedCycleGAN: Asymmetric Style Transfer for Applying and Removing Makeup," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp40-48, 2018.

[8] Bing Li, Yuanlue Zhu, Yitong Wang, Chia-Wen Lin, Bernard Ghanem, Linlin Shen, "AniGAN: Style-Guided Generative Adversarial Networks for Unsupervised Anime Face Generation," IEEE Transactions on Multimedia, vol. 24, pp4077-4091, 2022.

[9] Yunjey Choi, Minje Choi, Munyoung Kim, Jung-Woo Ha, Sunghun Kim, Jaegul Choo, et al, "StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp8789-8797, 2018.

[10] Guilin Liu, Fitsum A. Reda, Kevin J. Shih, Ting-Chun Wang, Andrew Tao, Bryan Catanzaro, "Image Inpainting for Irregular Holes Using Partial Convolutions," Proceedings of the European Conference on Computer Vision (ECCV), pp85-100, 2018.

[11] Leon A. Gatys, Alexander S. Ecker, Matthias Bethge, "Image Style Transfer Using Convolutional Neural Networks," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp2414-2423, 2016.

[12] Justin Johnson, Alexandre Alahi, Li Fei-Fei, "Perceptual Losses for Real-time Style Transfer and Super-resolution," Computer Vision–ECCV 2016: 14th European Conference, pp694-711, 2016.

[13] Dmitry Ulyanov, Vadim Lebedev, Andrea Vedaldi, Victor Lempitsky, "Texture Networks: Feed-forward Synthesis of Textures and Stylized Images," 33rd International Conference on Machine Learning, pp 2027-2041, 2016.

[14] Dongdong Chen, Lu Yuan, Jing Liao, Nenghai Yu, Gang Hua, "Stylebank: An Explicit Representation for Neural Image Style Transfer," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp1897-1906, 2017.

[15] Zhou Chang, "Study on Image Style Transfer Algorithm Based on Deep Learning," Yangzhou University, 2020.

[16] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros, "Image-to-image Translation with Conditional Adversarial Networks," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp1125-1134, 2017.

[17] Jun-Yan Zhu, Taesung Park, Phillip Isola, Alexei A. Efros, "Unpaired Image-to-image Translation Using Cycle-consistent Adversarial Networks," Proceedings of the IEEE International Conference on Computer Vision, pp2223-2232, 2017.

[18] Sun Tianpeng, Zhou Ningning, Huang Guofang, "New GAN-Based Partial Realistic Anime Image Style Transfer," Computer Engineering and Applications, vol. 58, no. 14, pp167-176, 2022.

[19] Mirza M, Osindero S, "Conditional Generative Adversarial Nets," arXiv preprint arXiv:1411.1784, 2014.

[20] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, et al, "Imagenet Large Scale Visual Recognition Challenge," International Journal of Computer Vision, vol. 115, no. 3, pp211-252, 2015.

[21] Dmitry Ulyanov, Andrea Vedaldi, Victor Lempitsky, "Instance Normalization: The Missing Ingredient for Fast Stylization," arXiv preprint arXiv:1607.08022, 2016.

[22] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, et al, "Microsoft Coco: Common Objects in Context," European Conference on Computer Vision, vol. 8693 ,pp740-755, 2014.

[23] Ren Guo-Xi, "Research on a Convolutional Neural Network Method for Modulation Waveform Classification," IAENG International Journal of Computer Science, vol. 50, no.3, pp875-882, 2023.

[24] Kang Tan, Linna Li, Qiongdan Huang, "Image Manipulation Detection Using the Attention Mechanism and Faster R-cnn," IAENG International Journal of Computer Science, vol. 50, no.4, pp1261-1268, 2023.

[25] Xizhou Yue, Guoyu Li, Pengyun Zhang, Qifeng Sun, Ning Chen, and Peiying Zhang, "A Method Based on Generative Adversarial Networks for Completion of Blank Bands in Electric Logging Images," Engineering Letters, vol. 32, no. 12, pp2371-2377, 2024

MA Chi, he received his Ph.D. in Dongling School of Economics and Management at the University of Science and Technology Beijing. Since 2010, he has been working as Associate Professor in School of Software Engineering at University of Science and Technology Liaoning. From 2019 to now, he has worked in the school of computer science and engineering at Huizhou University. His research interests include pattern recognition and data mining.

Wang Shaofan, he graduated from Zaozhuang University. Since 2019, he has been studying for master degree in the field of computer science and technology at School of Computer Science and Software Engineering, University of Science and Technology Liaoning. Since 2023, he has worked as an engineer at the the Derkee Information Co., Ltd. His research direction is deep learning and data mining.

Hu Hui, she received her Master's degree in computer applied technology with Nanjing Tech University, Nanjing, China. Since 2001, she has been working as lecturer in department of computing and information systems at Yancheng Institute of Technology. From 2011, she has worked in the school of computer science and engineering at Huizhou university. Her research interests include artificial intelligence, computer vision, image processing, pattern recognition.