AI Digital Anime Style Generation Algorithm Based on Adversarial Generative Network

Xiaojun Tan
Chengdu Neusoft University
China
yishu2009@163.com

Abstract

Digital computer technology is changing the development of animation image processing and significantly improving the efficiency of animation creation. Among them, traditional artificial animation faces problems such as low efficiency and poor quality in image drawing. Therefore, a research proposes a digital animation style generation model based on artificial intelligence. First, the animation style generation model is constructed based on the generative adversarial network. Considering the gradient vanishing problem, the residual network is introduced to perfect the model, and the attention mechanism is introduced to correct the image deviation problem. In the two scenarios of contour feature extraction and comic style transfer, the image loss of the research model is 0.012 and 0.038 respectively, which is better than similar models. In addition, in the comparison of animation style conversion quality, the research model handles the details of animation images better, and its peak signal-to-noise ratio is 23.05db, which is better than similar models. It can be seen that the research technology is superior to similar technologies in the field of animation creation and has good application effects. The research content will provide a technical reference for intelligent animation creation.

CCS Concepts

• **Computing methodologies**; • **Artificial intelligence**; • **Computer vision**;

Keywords

Artificial intelligence, Generative adversarial network, Residual network, Animation, Attention mechanism

ACM Reference Format:
Xiaojun Tan. 2024. AI Digital Anime Style Generation Algorithm Based on Adversarial Generative Network. In *9th International Conference on Cyber Security and Information Engineering (ICCSIE 2024), September 15–17, 2024, Kuala Lumpur, Malaysia.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3689236.3689257

ICCSIE 2024, September 15–17, 2024, Kuala Lumpur, Malaysia

ACM ISBN 979-8-4007-1813-7/24/09
https://doi.org/10.1145/3689236.3689257

1 Introduction

As a type of computer information processing technology, image processing technology, combined with artificial intelligence and other technologies, is changing the development of animation creation in recent years. Among them, animation, as a form of visual art expression, is different from the realistic style of scenery and content. When drawing anime, creators often consider the scene settings of characters and the artistic expression of anime. Therefore, creating anime is very time-consuming and labor-intensive [1]. At present, computer image processing technology based on artificial intelligence (AI) is quietly changing the creative thinking of anime. Through intelligent AI technology, it can generate anime style images, quickly color anime characters, and synthesize anime videos [2]. Currently, animation style image transfer technology is a key focus in the field of computer AI animation creation. For example, Goodfellow et al. proposed generative adversarial networks (GAN) in 2014. As an AI intelligent technology in the image field, it can generate new information based on distributed probability data and has important applications in text, image, video and other fields [3]. In addition, Tango K and other scholars optimized and improved the traditional GAN model. By integrating it with the convolutional network, the GAN model further enhanced the processing of image feature data .Tango K et al. conducted research in the field of animation images, and the generation technology of image to image style has attracted attention in the field of anime cosplay. Therefore, in order to improve character style and clothing stylization, research and design image generation technology based on adversarial generative networks, and achieve image processing and generation through data collection and processing. Compared with traditional manual design, research technology has diversified styles and significantly improved production efficiency [4]. Hinami R et al. conducted research on existing image processing techniques and found that the combination of artificial intelligence and image processing technology can significantly improve the translation innovation effect of comics. In this regard, a deep learning technique was used to construct an image processing model, and secondly, the model's image processing performance was improved by extracting and training the image context. And apply it to the process of anime translation and production, research technology has better efficiency compared to traditional technology, and the innovative effect is more outstanding [5]. It can be seen that AI technology has important applications in the fields of image and video. Therefore, to improve the problem of low efficiency of traditional artificial animation creation, this study proposes an improved AI animation style transfer technology based on GAN, which meets the requirements of animation creation by optimizing the parameters of the GAN model. The innovation of the research

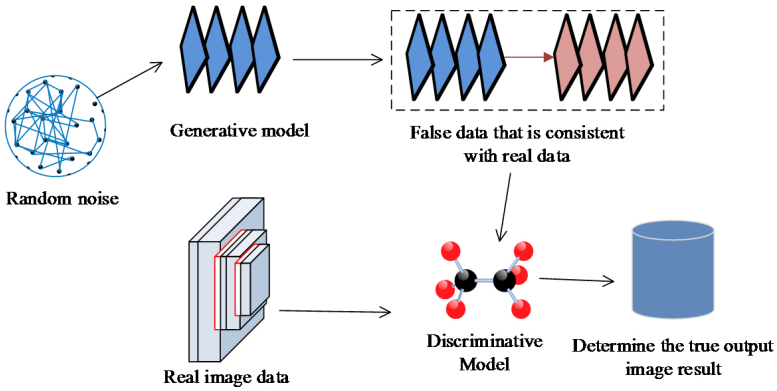

Figure 1: GAN model structure diagram

is to realize the generation of animation style images based on the advanced GAN network and improve the efficiency of animation creation. Secondly, the residual network is added to improve the model to ensure the quality of animation images. The research content will provide a reference for the application of AI technology in the animation industry.

2 Construction of An AI-based Animation Style Creation Generation Model

2.1 Construction of Anime Style Generation Model Based on GAN

At present, AI technology has gradually matured and has important applications in the field of animation creation. In order to improve the low efficiency of traditional artificial animation creation, this study proposes an intelligent animation style creation generation technology based on the GAN model, which is used for animation creators to generate animation style images. Among them, the GAN model mainly includes two important parts, namely the generation model, represented by G, and the discriminant model, D represented by [6]. The schematic diagram of the GAN model structure is shown in Figure 1.

In animation creation, the output animation objective function of the GAN model is defined as $\min_G \max_D(D, G)$, then the expression of the animation objective function generated by the GAN model is shown in formula (1).

$$\begin{aligned}\min_G \max_D (D, G) = E_{x-pd}\left[\log\left(D(x)\right)\right] \\ +E_{z-pz}[\log\left(1 - D\left(G\left(z\right)\right)\right)]\end{aligned} \tag{1}$$

In formula (1), p_d represents the probability distribution of real data in the GAN model, represents p_z the probability distribution obtained E from noise- generated data, z represents the maximum likelihood estimate, and x represents the input real image data. In the GAN model, real data and random noise are z input into the GAN model together, where the noise z enters the generator G to obtain a distribution close to the real data p_z. At the same time, the discriminator D will p_z judge the data. When the output data is real data x, the result output is 0. If the input data is generated $G(z)$, the result output is 0. In actual animation creation, in the GAN model, according to the black-and-white animation images and random noise provided by the creator, the generator G will generate color images corresponding to the required style according to the requirements [7]. The discriminator D is trained based on the generated images and the original color images to identify the real images in the generated images. The two networks are continuously optimized and trained in an adversarial form, and the image style and color transfer creation is realized through the image color mapping relationship [8]. In GAN model training, in order to improve the animation generation effect, it is usually required to maximize the discriminator function value as much as possible and the generator function value be minimized as much as possible. According to this rule, its goal is as shown in formula (2) [9].

$$G^* = \arg\min_G \max_D L_{cGAN}(G, D) \tag{2}$$

In formula (2), $L_{cGAN}(G, D)$ represents the loss function. Next, the discriminator x observes the input real image data, and the loss function is shown in formula (3).

$$\begin{aligned}L_{cGAN}(G, D) = E_{y-pd(y)}\left[\log D\left(y\right)\right] \\ +E_{x-pd(x),z-pz(z)}\left[\log(1 - D\left(G(x, z)\right)\right]\end{aligned} \tag{3}$$

In formula (3), $G(x, z)$ represents z the matching relationship between the noise z output images. When there is no noise z, the GAN model can still determine the matching relationship from the real image data x to obtain the output of image y, but this output belongs to deterministic output and cannot be matched with the distribution outside the delta function. Considering the impact of this condition on GAN training, in the generative model, the real image data x and Gaussian noise are used as the input of the generator, and the noise is provided after dropout processing [10]. In this way, the entropy generated by the model can be obtained in the noise training output, improving the model training effect.

2.2 Construction of Anime Style Generation Model Based on Improved GAN

In GNA model training, GNA training is affected by the number of model layers. The increase in the number of layers increases the number of parameters, causing the model to face gradient vanishing and overfitting problems, affecting the quality of image details. In order to optimize the gradient vanishing problem caused by the

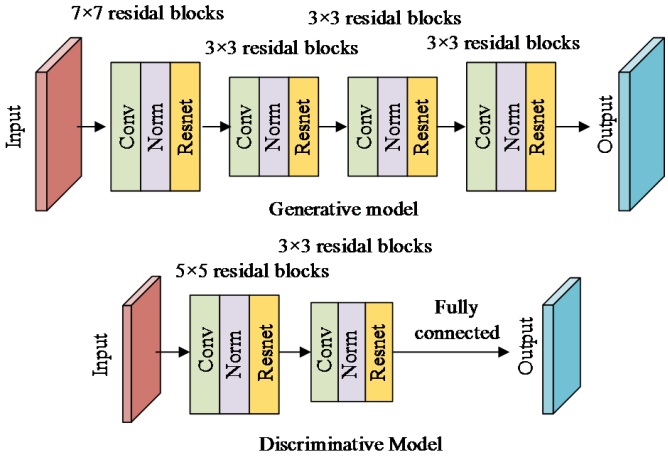

Figure 2: Improved GAN model network structure

increase in the number of network depths, the residual network (ResNet) is introduced for model optimization [11]. The residual network no longer fits the expected feature mapping relationship by stacking multiple nonlinear layers, but explicitly performs residual mapping through these layers [12]. The expected mapping is defined as $H(x)$, as calculated by formula (4).

$$F(x) = H(x_1) - x_1 \tag{4}$$

In formula (4), x_1 denotes the input value. The output value of the residual network is shown in formula (5).

$$y = F(x, \{W_i\}) + W_s x \tag{5}$$

In formula (5), F is an input value regularization function. W_s is x_1 the weight matrix between. $\{W_i\}$ is a string of weight matrices. In order to optimize the vanishing gradient problem of the GAN model, a residual module is added to the generator and the discriminator. The improved GAN model network structure is shown in Figure 2.

According to the structure in Figure 2, the generator uses a 7×7 residual block to extract image feature data, and then uses a 3×3 residual block to reduce the dimension of the data. Finally, the 7×7 and 3×3 deconvolution blocks are used to generate the anime style. In addition, a RELU activation function is added to each network layer in the generator to convert the linear output into a nonlinear output. In the discriminator, a 5×5 residual block is used for encoding and classification to obtain basic local features [13]. At the same time, it is returned through a 3×3 residual block and the output image result is obtained by full connection. In the discriminator, sigmoid and RELU activation functions are mainly used for data processing. Next, in order to ensure that the generated animation meets the high quality requirements, it is also necessary to analyze the loss function in the network. The study uses the anime-style CartoonGan network to reflect the true loss of retaining the original image data content, as shown in formula (6).

$$L(G, D) = L_{cGAN}(G, D) + \omega L_{con}(G, D) \tag{6}$$

In formula (6), the loss function consists of two parts, where $L_{cGAN}(G, D)$ is the loss function of the ordinary GAN network, $L_{con}(G, D)$ represents the loss function of the retained image real content, and ω represents the edge processing parameter. In $L_{con}(G, D)$ the loss function, L_2 the loss is replaced by L_1 the loss, which belongs to the regularization process [14]. In GAN training, although the residual network corrects most of the coloring problems, the problem of detail loss still occurs when performing style coloring on images through training. In order to further correct the above problems of the GAN model, a convolutional attention mechanism is added to it for modification. The convolutional attention mechanism will be used for feature preprocessing in the feature extraction stage, so that different training parameters for different parts can be obtained according to the training, which will further enhance the GAN model's emphasis on different parts [15]. The attention mechanism calculation in the feature map channel is shown in formula (7).

$$\begin{aligned} M_c(F_r) &= \sigma(MLP(AvgPool(F_r)) + MLP(MaxPool(F_r))) = \\ &\sigma(W_1(W_0(F_{r\,avg}^c)) + W_1(W_0(F_{r\,max}^c))) \end{aligned} \tag{7}$$

In formula (7), F represents the input feature map, $MaxPool(F_r)$ represents the maximum pooling, $AvgPool(F_r)$ represents the average pooling, MLP represents the multi-layer perception mechanism, W_0 and W_1 represents two fully connected weights, and σ represents the activation function. In this process, the maximum pooling and the average pooling are relative to the feature map, so the obtained are $1\times c$ feature vectors. In MLP the perception layer, the two layers share information with each other. MLP The results of the perception layer will be added and processed by σ the function to obtain the attention feature map. Finally, the result will be multiplied with the input feature to obtain the final channel attention result.

3 Analysis of model application effect

3.1 Model Experimental Environment

To test the actual application effect of the animation generation model proposed by the institute, the study will carry out specific experiments to validate the model's performance. The experimental system is WINDOWS 10 Professional Edition, the running memory is 64G, the graphics card is NVIDIA P1000 model, the processor is INTEL i9, and the running environment is Python 3.5 + Tensorflow-gpu 1.13.1. The image animation generation adversarial network (Cartooon-Generative adversarial nets, CartoonGAN) and the background generation adversarial network (Scenery-Generative adversarial nets, SceneryGAN) are introduced as test benchmarks. The training data set randomly crawls the Safenooru animation website and various portal website materials through crawler technology, retaining a total of 28,000 animation image data suitable for training, so that all animation images are 255×255 in size. In addition, the peak signal-to-noise ratio (PSNR) and image loss (Image loss) are introduced as evaluation indicators in the experiment.

3.2 Technical Effect Analysis

The experiment first compares the effects of different models on the extraction of image edge contour features. Cat image data that has been converted into a hand-painted style is selected for the experiment. The comparison of the extraction effects of different models on the cat contour edges is shown in Figure 3.

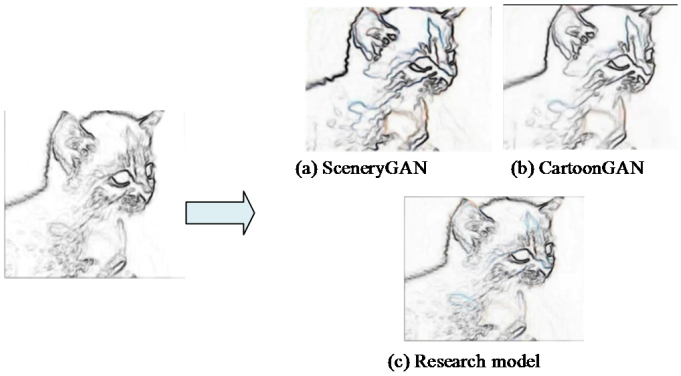

Figure 3: Comparison of contour extraction results for different models

Figure 4: Comparison of image migration effects for anime styles of different models

In Figure 3, the left side shows a hand-drawn style cartoon image processed by the OpenCV image tool, which is taken as the image input to compare the effects of various models on the extraction of cat edge contour details. Figures 3 (a) to 3 (c) are SceneryGAN, CartoonGAN and the model proposed by the research institute. Given these test results, the SceneryGAN is not clear in processing the details of the cat's edge contour, especially when the image is enlarged, it can be seen that some of the cat's hair details have been removed. In addition, the cat's facial contour is distorted, but the overall contour is well preserved. The CartoonGAN model handles the cat's edge contour details better than the SceneryGAN, retaining some of the cat's edge hair, and the extraction of the cat's facial contour is smoother, but the cat's contour still has edge detail distortion, but the contour distortion is smaller than that of the SceneryGAN. The research method owns the greatest edge detail processing, which can well retain most of the contour details of the original image, especially the edge hair can still be clearly seen, and there is no obvious distortion problem in the cat's facial details, and the overall contour extraction of the cat is the best. Next, we compare the animation style transfer effects of different models, as shown in Figure 4.

Figure 4 compares the effects of different models in transferring anime-style images, with the original real image on the left. In the image processing of Person 1 holding flowers, the three models have obvious differences in style processing. The SceneryGAN model has obvious color cast problems in the processing of characters, such as unreasonable red blocks on the face of the character, and serious loss of details of the facial features, which does not

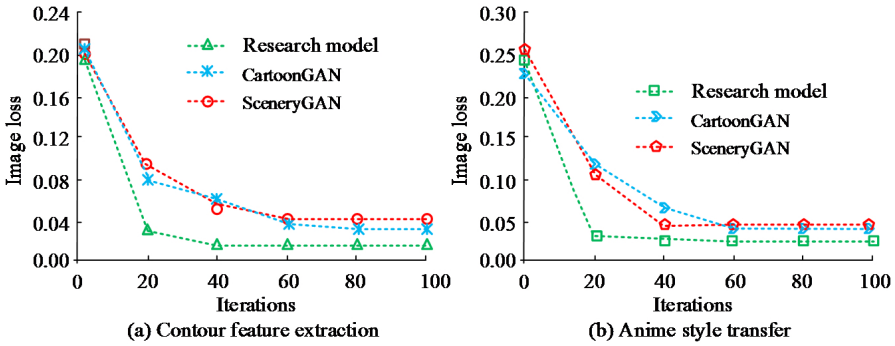

Figure 5: Comparison of image loss results for different models

have the characteristics of anime images. The CartoonGAN model handles the details of the image of the person holding flowers more reasonably, without color deviation problems, but the details of the face and flowers are lost more. The best performance is the research model, which retains more overall details and has no color deviation problems. The processing of Person 2 is basically the same as that of Person 1. Both the SceneryGAN model and the CartoonGAN model have problems such as detail loss and color deviation. In the third cat picture, more detailed contours test the model's extraction of image features and color processing. The SceneryGAN model and the CartoonGAN model have basically the same processing of cat details, but the SceneryGAN model has obvious color deviation problems. For example, the patch on the cat's forehead is already light yellow. The CartoonGAN model processes the cat's color slightly better, but the overall tone of the cat is cold, and the facial contour details are lost. The best performance is the research model, which has a warmer style and is closer to the anime style. It retains more details of the cat's ears and face, and the overall processing is better. Finally, there is a comparison of static buildings, where SceneryGAN and CartoonGAN show significant loss of details, but CartoonGAN still preserves the tree outline. However, both have a white tint in image processing, which does not conform to the animation style. Compared to other models, research models have better details and color accuracy, while retaining animation colors, which is more in line with the requirements of animation creation. Next, the image loss of different models in the two processes of contour feature extraction and anime style transfer is compared, as shown in Figure 5.

Figure 5 (a) is a comparison of the results of the contour feature extraction process, where the SceneryGAN model converges after 60 iterations, and the image loss is 0.062 at this time. The CartoonGAN model converges after 80 iterations, and the image loss is 0.056 at this time. The research model performs the best, which converges after 40 iterations, and the image loss is 0.012 at this time. Overall, the research model has higher accuracy in image contour feature extraction, and the model converges faster. Figure 5 (b) is a comparison of the results of the animation style conversion process. The research model has the optimal performance, which can converge after 20 iterations with the image loss of 0.038 at this time. The training effects of the SceneryGAN and the CartoonGAN are similar, but the SceneryGAN can converge faster, converge after 40 iterations, and the image loss is 0.051 at this time, while the CartoonGAN converges after 60 iterations, and the image loss is 0.050 at this time. Figure 6 is a comparison of image quality processing results of different models.

Figure 6 (a) shows the comparison of the results of the contour feature extraction process. The best image processing effect is the research model, which converges after 40 iterations and the image PSNR value is 24.05db, which is better than others. This indicates that the research model is superior at processing image contour details and ensures the quality of the image. The second best performance is the CartoonGAN model, which also converges after 40 iterations and the image PSNR value is 21.05db. Because the CartoonGAN loses the details of the image edge contour processing, the PSNR value is lower than that of the research model. The worst performance is the SceneryGAN model, with an image PSNR value of 18.08db at convergence. At the same time, the image quality processing effect of the model in the animation style transfer process is compared, as shown in Figure 6 (b). From the results, the research model still performs better, with an image PSNR value of 23.05db during the convergence period, while the CartoonGAN model and SceneryGAN model are 20.05db and 16.08db respectively.

4 Conclusion

AI and computer digital image processing technology are changing the efficiency and quality of animation creation. To improve the efficiency of animation image generation, an intelligent animation style transfer generation technology based on GAN network is proposed. Considering the problem of color detail loss in image processing, ResNet and convolutional attention mechanism are introduced for optimization. By comparing the migration effects of various anime style images, the research model can well retain the facial details of the characters, and the color is closer to the style of anime images, and there is no image color deviation. In the image processing of architectural styles, research models can effectively preserve details and perform better in static image processing. In the image loss comparison, the image loss of the research model in contour feature extraction is 0.012, while SceneryGAN and CartoonGAN are 0.062 and 0.056. In addition, the image loss of the research model in the process of anime style conversion is 0.038, which is better than other models. In addition, the processing effects of different models on anime images are compared. In the

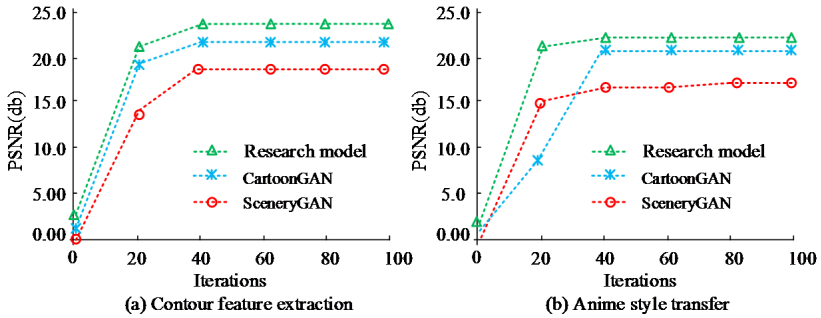

Figure 6: Comparison of image quality processing results for different models

process of style migration of anime images, the image PSNR values of the research model, SceneryGAN, and CartoonGAN are 23.05db, 20.05db, and 16.08db respectively. From this, it can be seen that the animation style generation technology proposed by the research institute has significant advantages in image processing quality, color processing, and detail optimization compared to similar technologies, and is more in line with the requirements of the animation creation field. However, although residual networks were added to alleviate the gradient vanishing problem in GAN models in the research, the processing of more complex anime images still faces high computational complexity. In the future, efforts need to be made to simplify the model and reduce its computational load.

References

[1] White D, Katsuno H. Toward an affective sense of life: artificial intelligence, animacy, and amusement at a robot pet memorial service in Japan. Cultural Anthropology, 2021, 36(2): 222–251.

[2] Su H, Niu J, Liu X, Wan J. Mangagan: Unpaired photo-to-manga translation based on the methodology of manga drawing. Proceedings of the AAAI Conference on Artificial Intelligence. 2021, 35(3): 2611-2619.

[3] Oshiba J, Iwata M, Kise K. Face image generation of anime characters using an advanced first order motion model with facial landmarks. IEICE TRANSACTIONS on Information and Systems, 2023, 106(1): 22-30.

[4] Tango K, Katsurai M, Maki H. Anime-to-real clothing: Cosplay costume generation via image-to-image translation. Multimedia Tools and Applications, 2022, 81(20): 29505-29523.

[5] Aoyama R, Ng R. Artificial flavors: nostalgia and the shifting landscapes of production in Sino-Japanese animation. Cultural Studies, 2024, 38(2): 245-272.

[6] Hinami R, Ishiwatari S, Yasuda K. Towards fully automated manga translation. Proceedings of the AAAI Conference on Artificial Intelligence. 2021, 35(14): 12998-13008.

[7] Ambrosetti N. Fighting with Rotating Blades, Boomerangs, and Crushing Punches: A History of Mecha from a Robotics Point of View. Foundations of Science, 2024, 29(1): 59-85.

[8] Zhao J, Xu S, Chandrasegaran S. Chartstory: Automated partitioning, layout, and captioning of charts into comic-style narratives. IEEE transactions on visualization and computer graphics, 2021, 29(2): 1384-1399.

[9] Peng Y. Bilateral Filtering Based Image Cartoon Stylization Design. Academic Journal of Computing & Information Science, 2022, 5(10): 70-77.

[10] Zhao Y, Ren D, Chen Y, Jia W. Cartoon image processing: a survey. International Journal of Computer Vision, 2022, 130(11): 2733-2769.

[11] Plutino A, Barricelli B R, Casiraghi E, *et al.* Scoping review on automatic color equalization algorithm. Journal of Electronic Imaging, 2021, 30(2): 020901-020901.

[12] Mataram S, Purwasito A, Subiyantoro S, *et al.* Developing Of A Comic Digital Gallery To Improve The Potential Of Visual Communication Design Students. Journal of Positive School Psychology, 2022, 6(10): 3765-3783.

[13] Sharma S, Verma K, Hardaha P. Implementation of artificial intelligence in agriculture. Journal of Computational and Cognitive Engineering, 2023, 2(2): 155-162.

[14] Pally R J, Samadi S. Application of image processing and convolutional neural networks for flood image classification and semantic segmentation. Environmental modelling & software, 2022, 148(2): 21-25.

[15] Yu J, Jin L, Chen J, Xiao Y, Tian Z, Lan X. Deep semantic space guided multi-scale neural style transfer. Multimedia tools and applications, 2022, 81(3): 3915-3938.