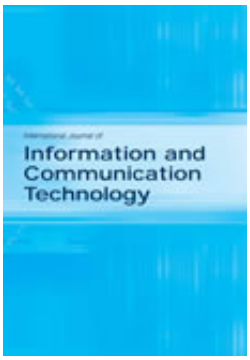
INDERSCIENCE PUBLISHERS
Linking academia, business and industry through research



International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

The style transfer model of illustration images based on multi-scale CycleGAN

Yanran Liang, Yumeng Yan

Article History:

Received:	20 December 2024
Last revised:	12 February 2025
Accepted:	13 February 2025
Published online:	15 April 2025

The style transfer model of illustration images based on multi-scale CycleGAN

Yanran Liang and Yumeng Yan*

College of Fine Arts Education,
Guangxi Arts University,
Nanning, 530000, Guangxi, China
Email: liangyanran@163.com
Email: yympeach@126.com
*Corresponding author

Abstract: Traditional image style transfer methods cannot preserve the content and structural features of the original image while maintaining a specific style. To preserve the semantic information of the original image, a multi-scale cycle-consistency generative adversarial network model is developed. This model can enable innovative style transformations while maintaining the original artistic characteristics of illustrations. This model can better capture and integrate detailed features of different styles by performing style transfer at different resolution levels. The results showed that the proposed model improved the inception score by 1.755 and 0.122 respectively compared to the other two methods, indicating a significant improvement in image generation quality and superiority in image generation. When the low-level texture feature loss, adversarial loss, and high-level concept feature loss were removed, the Frechet inception distance value significantly increased from 73.72 to 102.28, an increase of approximately 38.74%, emphasising the role of these components in the model. The model proposed in this study achieves diverse style transfer and can maintain high image quality when generating stylised images, providing artists and designers with greater creative inspiration and choice space.

Keywords: multi-scale CycleGAN; generative adversarial networks; illustration images; style transfer; classification.

Reference to this paper should be made as follows: Liang, Y. and Yan, Y. (2025) 'The style transfer model of illustration images based on multi-scale CycleGAN', *Int. J. Information and Communication Technology*, Vol. 26, No. 7, pp.1–16.

Biographical notes: Yanran Liang obtained her Doctoral degree from Silpakorn University in Thailand. Her main research directions are design and illustration.

Yumeng Yan obtained her Doctoral degree from Silpakorn University in Thailand. She is currently working at Guangxi Arts University in China as a senior arts and crafts master. Her main research directions are visual communication and decorative patterns.

1 Introduction

In the digital art and image processing, style transfer technology is an innovative approach that allows artists and designers to give their original image content a new visual style while preserving it. This technology not only enriches the forms of artistic expression, but also provides new possibilities for image editing and visual effects production (Chen et al., 2024; Yan et al., 2024). As the digital media and online culture develop, the role of illustration art in visual communication and cultural expression is increasingly prominent. Through style transfer techniques, more possibilities can be provided for illustration creation, allowing artists to explore and experiment with different visual languages while maintaining their personal style. Although various style transfer methods have been proposed, there are still certain limitations when dealing with specific types of images such as illustrations, comics, etc. (Han et al., 2024; Wang et al., 2024). For example, generative adversarial networks (GANs) introduced a cyclic consistency loss function on the basis of unidirectional mapping, which to some extent avoids model collapse through bidirectional mapping. However, instability may sometimes occur during training (Azni et al., 2023). Moreover, traditional GANs may encounter problems such as loss of details or inconsistent styles when processing complex images (Nammee, 2023). As a cutting-edge image processing technology, cycle-consistent generative adversarial network (CycleGAN) can achieve high-quality image style transfer without the need for a large amount of paired data, which makes it highly applicable in artistic creation and commercial design.

Image style transfer and editing techniques have broad application prospects in practical applications, including game development, film production, virtual reality, and other fields. Satchidanandam et al. (2023) combined the subjective loss algorithm of deep neural networks with semantic segmentation technology to enhance the aesthetic correctness of style transfer, and integrated it into GAN to achieve automatic segmentation for precise understanding of image meaning. Experiments showed that this method significantly outperformed traditional methods in terms of visual accuracy. Wang (2023) used neural networks to extract style and content, and achieved ethnic clothing style transfer through image reconstruction techniques. The shoulder affine transformation in colour space constrained the transformation of input and output images, effectively suppressing image distortion. Gao et al. (2021) proposed a wallpaper texture generation and style transfer framework grounded on multi-label semantics and GAN. This method evaluated the authenticity of generated wallpapers and the degree to which they conform to specified attributes by training a perception model, and generated wallpaper images with specific styles using multi-label semantics as conditional variables. The experiment findings confirmed that this method could generate wallpaper textures that conform to human aesthetics and have artistic features. Richter et al. (2022) developed a new image patch sampling strategy to address the differences in scene layout distribution in existing datasets, and introduced architecture improvements for multiple deep network modules. The experiment outcomes denoted that compared with existing image to image translation methods and other baselines, this method has made significant progress in stability and practicality. Durrant (2022) developed a deep learning model, designated Prot2Prot, which is capable of rapidly emulating authentic visualisation styles and facilitating the generation of molecular representations that are readily comprehensible. Compared with traditional 3D graphics programs, Prot2Prot could create images in a short amount of time and even run efficiently in web browsers.

Chen et al. (2024) put forth a system for transforming facial photographs into portraits with a distinctive charcoal sketch style. The system employed CycleGAN to generate paired examples of Pix2Pix, thereby enabling the conversion of photos into comics. The findings indicated that the images generated by the system could effectively reproduce the comedic style, especially in the facial area. To raise the quality of generated images, Yan et al. (2024) proposed an SAR image ship wake data enhancement method based on improved CycleGAN. To resolve the issue of incomplete data in the generated images at the microscopic level, a least squares loss was utilised. Moreover, a convolutional block attention module was integrated into the decoder of the generator with the objective of improving the quality of the generated images. Sugiyama and Aikawa (2024) proposed a method for detecting defects by utilising the differences between the pseudo images generated by CycleGAN and the original images. Compared with traditional binary detection methods, this method could detect defects independently of the shooting environment, greatly reducing the risk of ignoring defects.

In summary, the existing models are limited in practical applications and difficult to widely apply to various style transfer tasks. The style transfer model based on GAN is prone to pattern collapse and overfitting, and the generator may 'remember' a few style images, resulting in copying these images during generation. Despite the fact that numerous models have been developed to achieve multi-domain style transfer, the domain of style transfer remains constrained, which makes it challenging to achieve more detailed transformations. In response to the above issues, an innovative illustration image style transfer model based on multi-scale CycleGAN is proposed. By using an improved generator to transform illustration images from one domain to another, the discriminator determines the authenticity of the illustration images. Then, by introducing a cyclic consistency loss function, it ensures that the transformed illustration images remain consistent in content without relying on paired data.

The main contributions of this research include:

1 a multi-scale CycleGan-based illustration image style transfer model is proposed, which can achieve high-quality style transfer while maintaining the image content structure

2 by introducing multi-scale generator and discriminator, the training stability and image generation quality of the model are significantly improved

3 the superior performance of the model on multiple datasets is verified through experiments, especially the significant improvement in image similarity and style retention.

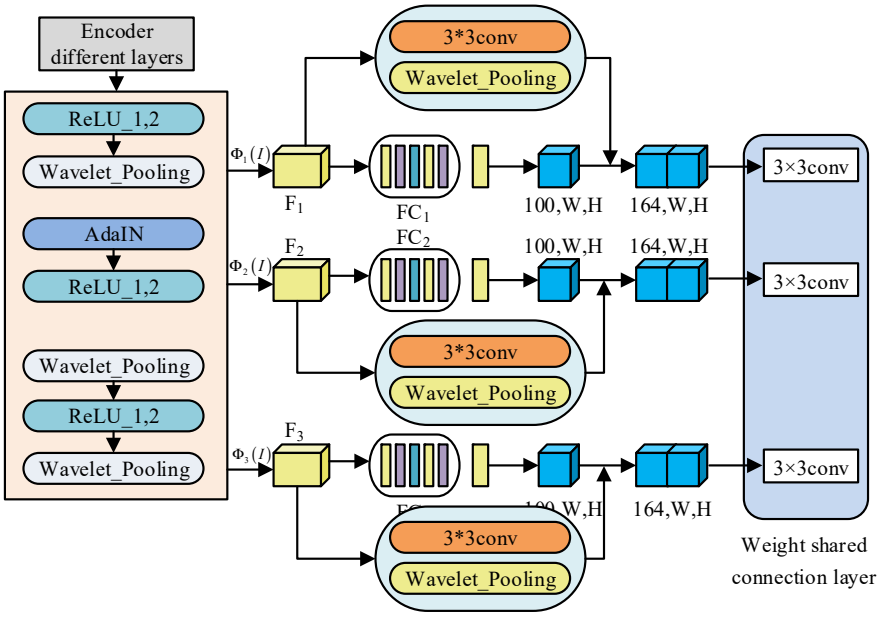
2 Methods and materials

2.1 Feature extraction of illustration images

In transferring illustration style information, traditional single channel feature transformation paths usually only focus on single dimensional features such as colour, texture, or shape, while ignoring the interaction between these features and the overall artistic effect. However, in the content and style encoding stage, if effective semantic and style associations are not established, fusion errors may occur in the decoding stage,

which cannot capture the deep style features in the illustration, such as complex textures and details, resulting in inaccurate transmission of style features (Yan et al., 2024). The domain sense indicator refers to the quantitative index used to measure whether the image generated by the model in the target domain (TD) (such as different illustration styles) conforms to the semantic and artistic characteristics of the domain in the task of cross-domain image style transfer. It can help the network capture the attributes of the domain from a given reference image, and then adaptively adjust the degree of stylisation and structural preservation based on these attributes. This design enables DSTN to transition between artistic style and photo realistic style, generating high-quality stylised results regardless of the target field. Therefore, the study introduces domain sensitive indicators into the network structure of illustration image feature extraction, as shown in Figure 1.

Figure 1 Illustration image feature extraction network structure (see online version for colours)



In the style information encoding stage, the feature maps output by each layer of the encoder are processed through a $3 * 3$ convolutional layer and wavelet pooling operation is introduced to form a channel set feature map. The purpose of this step is to combine style information with content features so that both the content and style of the image can be considered in subsequent processing. The feature map W_i^{DI} output by each layer of the encoder can be expressed by equation (1).

$$W_i^{DI} = \lambda\left(\left[FC_i\left(Gram\left(W_i\right)\right) \otimes W_i'\right]\right) \tag{1}$$

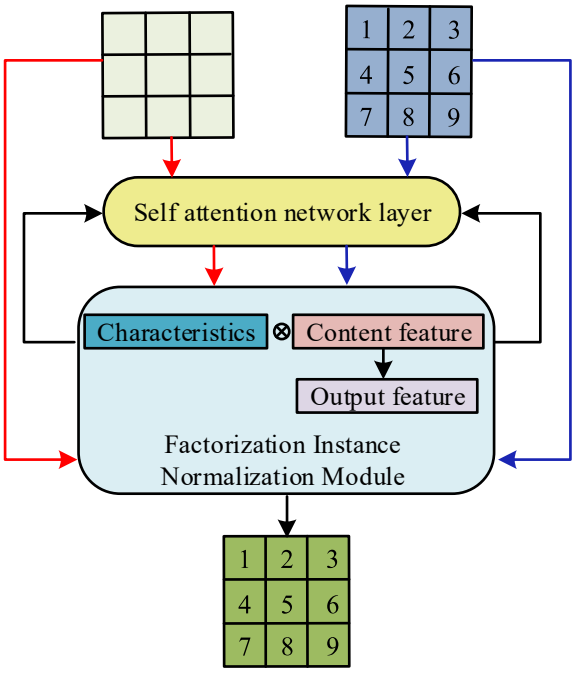
In equation (1), W_i refers to the feature map extracted from the i^{th} layer; W_i' denotes the feature map processed by the channel attention mechanism; $\otimes$ stands for channel connection operator, used to integrate information between channels; the fully connected layers FC and λ with weight sharing are used to further process these features. The style

of illustration images often exhibits distinct regional features, indicating that different parts of the image may display vastly different style information. To preserve this regional style feature during style transfer, the model used must be able to recognise and process the information of these style domains separately (Sugiyama and Aikawa, 2024; Zhao et al., 2021). This paper studies the weighted fusion of input features through the attention mechanism, so as to enhance the feature representation of important regions. The attention mechanism computes the weight A for each spatial position through two convolutional layers and an activation function, as shown in equation (2).

$$\begin{cases} A = \sigma\left(W_2 \bullet \mathrm{ReLU}\left(W_1 \bullet F\right)\right) \\ F_{fused} = F \odot A \end{cases} \tag{2}$$

In equation (2), W_1 is the weight matrix of the first convolution layer, and W_2 is the weight matrix of the second convolution layer. σ is the sigmoid activation function, which normalises the output to the [0, 1] range, representing the attention weight for each spatial location. $\odot$ means multiplication-by-element. The calculated attention weight A is used to weight the input feature F, and the final fusion feature F_{fused} is the product of the input feature and the attention weight, which emphasises the features of important regions.

Figure 2 Self-attention semantic feature matching channel (see online version for colours)



The model for transferring illustration styles should not only maintain the overall style (i.e., the unified style of the entire image), but also maintain the local style (i.e., the style of specific areas of the image). This requires the model to accurately process style features to ensure that each region in the final generated image can accurately display its

expected style. To achieve adaptive fusion of style features and content features, a self attention semantic feature matching channel grounded on the self-attention factorised instance normalisation (SAVIN) module is proposed in the study, as shown in Figure 2. This self attention mechanism enables the model to simultaneously consider global information when processing features, thereby more effectively understanding the overall context of the image (Yang et al., 2021; Gupta et al., 2019; Kim et al., 2023).

The study takes the content feature $\overline{W_d'}$ and style feature $\overline{W_s'}$ as inputs, and learns the normalisation parameters b_s and c_s by identifying the semantic correspondence and key features between $\overline{W_d'}$ and $\overline{W_s'}$. The description of the conversion process can be represented by equation (3).

$$\overline{W_d'}, \overline{W_s'} = c_s \otimes \left(a_{d\_s} \left(\overline{W_d}, \overline{W_s} \right) + a_{s\_d} \right) + b_s \tag{3}$$

In equation (3), a_{d_s} and a_{s_d} are used to capture the semantic correspondence between $\overline{W_d'}$ and $\overline{W_s'}$ before style matching. The calculation methods for b_s and c_s are shown in equation (4).

$$\begin{cases} c_s = ReLU\left(conv_{1\times1} \otimes SA_\gamma \left(\overline{W_c}, \overline{W_s} \right) \right) \\ b_s = ReLU\left(conv_{1\times1} \otimes SA_\beta \left(\overline{W_c}, \overline{W_s} \right) \right) \end{cases} \tag{4}$$

In equation (4), SA_γ and SA_β enhance feature representation through self attention mechanisms, thereby enabling the model to direct its focus towards the key elements of the image, thereby raising the quality and accuracy of style transfer. In this way, the model can not only learn the correspondence between content and style, but also effectively transfer style features to the content image while maintaining the content structure.

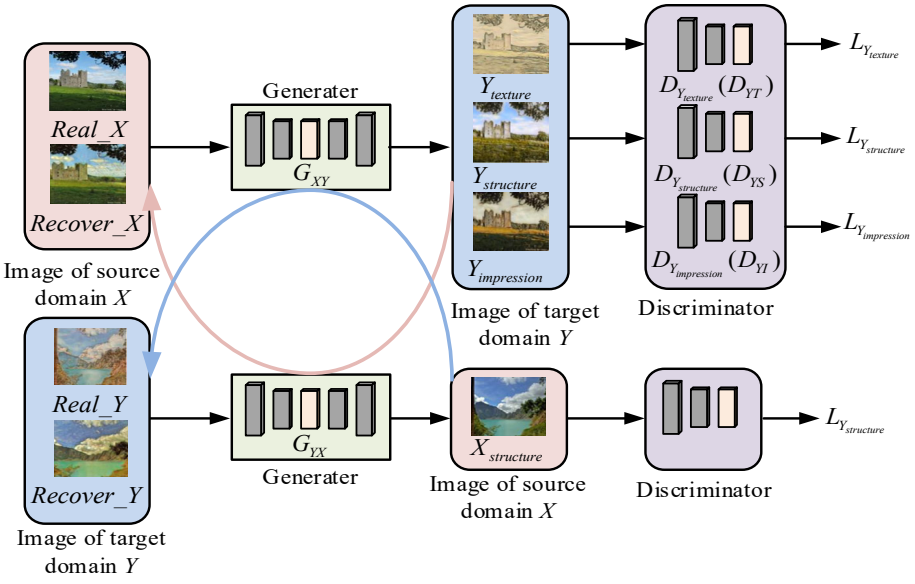
2.2 Construction of illustration style transfer model based on multi-scale CycleGAN

Traditional image transformation models typically require a large amount of paired training data, while CycleGAN only requires two sets of images from different domains, without one-to-one correspondence. By introducing cyclic consistency loss, CycleGAN can learn more stable and meaningful mapping relationships. However, traditional CycleGAN may lose some detail information when dealing with image style transfer. To address this issue, a multi-scale CycleGAN model is raised, which can better preserve the detailed information of images by learning their content and style features at different scales. The generator framework of multi-scale CycleGAN is denoted in Figure 3.

In Figure 3, *Real_X* and *Real_Y* respectively represent real image instances from the image distribution domain *X* and the illustration image distribution domain *Y*. G_{XY} is a generator network responsible for converting images from the *X* domain to the *Y* domain, while G_{YX} transfers images from the *Y* domain to the *X* domain. $Y_{structure}$ is an image generated by G_{XY} grounded on the probability distribution of the *Y* domain. $Y_{texture}$ is a single channel grayscale image obtained by applying function transformation to $Y_{structure}$, aimed at extracting texture features of the image. $Y_{impression}$ applies Gaussian blur technique to process $Y_{structure}$ and generate blurred images, aiming to capture the overall conceptual information of the image while maintaining smooth transitions of edges and

eliminating texture and details of the image. D_{YT}, D_{YS} and D_{YI} are three discriminative networks used to evaluate the underlying texture features, adversarial loss, and high-level conceptual features of images, respectively.

Figure 3 Schematic diagram of the generator structure (see online version for colours)



The effectiveness of network structure depends on the careful design of the loss function. A multi-scale adversarial loss function L was developed, which consists of three parts: the low-level texture feature loss $L_{Y_{texture}}$, adversarial loss $L_{Y_{structure}}$, and high-level conceptual feature loss $L_{Y_{impression}}$ of the image, corresponding to the fitting degree of the generated image in the TD at three different semantic levels. Texture loss focuses on the high-frequency texture features generated during the unidirectional mapping process from the source domain to the TD, without involving other factors such as colour or brightness of the image. The specific expression of the loss function is shown in equation (5).

$$L_{Y_{texture}} = E_{y\sim P_{data(y)}}\left[\ln D_{YT}(y)\right] + E_{x\sim P_{data(x)}}\left[\ln\left(1 - D_{YT}\left(K\left(G_{XY}(x)\right)\right)\right)\right] \tag{5}$$

In equation (5), x is the source domain image, y is the TD image, and $E_{x\sim P_{data(x)}}$ is the data distribution. K represents a function processing step that does not include a neural network. Similar to the loss of low-level texture features, the design of high-level concept loss focuses on the colour composition and surface features of illustration images. The specific expression for this loss is shown in equation (6).

$$L_{Y_{impression}} = E_{y\sim P_{data(y)}}\left[\ln D_{YI}(y)\right] + E_{x\sim P_{data(x)}}\left[\ln\left(1 - D_{YI}\left(V\left(G_{XY}(x)\right)\right)\right)\right] \tag{6}$$

In equation (6), V refers to the function that applies Gaussian blur processing to the generated image. The structural loss of the intermediate layer adopts the adversarial loss in GAN, and its specific loss expression is shown in equation (7).

$$L_{Y_{structure}} = E_{y\sim P_{data(y)}}\left[\ln D_{YS}(y)\right] + E_{x\sim P_{data(x)}}\left[\ln\left(1 - D_{YS}\left(G_{XY}(x)\right)\right)\right] \tag{7}$$

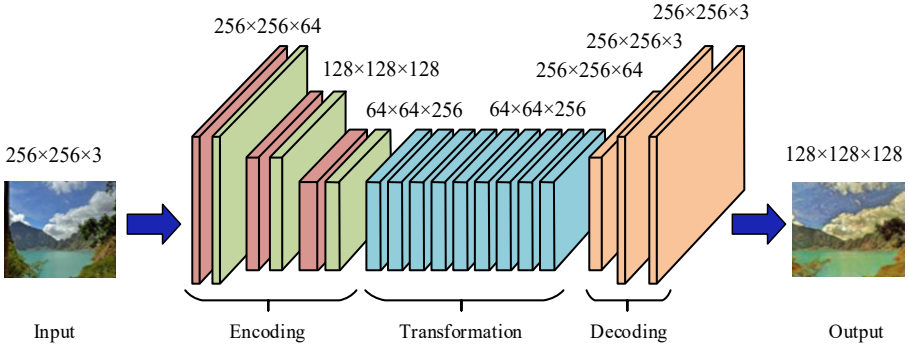
The goal of the research is to optimise the mapping process from the source domain to the TD while reducing the mapping strength from the TD to the source domain. Therefore, for images in the TD, the study adopted the multi-scale adversarial loss mentioned above to replace the adversarial loss in the traditional CycleGAN model, to guide the optimisation process of the network more strictly. An adaptive style weighting mechanism is also introduced to dynamically adjust the weight of style loss according to the complexity of style images to ensure that style features are not diluted. In the concrete implementation, the research uses the gradient information of style image to measure its complexity and dynamically adjusts the weight of style loss. The calculation formula is shown in equation (8).

$$L = \lambda^{(l)}_{texture} L_{Y_{texture}} + \lambda^{(l)}_{structure} L_{Y_{structure}} + \lambda^{(l)}_{impression} L_{Y_{impression}} \tag{8}$$

In equation (8), $\lambda^{(l)}_{texture}$, $\lambda^{(l)}_{structure}$ and $\lambda^{(l)}_{impression}$ are the weights of texture feature loss, adversarial loss and high-level concept feature loss at layer l, respectively. In each iteration, the gradient information and complexity of the style image are recalculated, the style loss weight is dynamically adjusted, and the generated image is optimised.

To enhance the generalisation ability of the network, it is necessary to consider the interaction between the extracted image features in the original network, which may make the network sensitive to external factors such as the dataset. The multi-scale CycleGAN introduces multi-scale generators, each responsible for converting images of different resolutions, as shown in Figure 4. These generators gradually convert low resolution images into high-resolution images through cascading, thereby achieving multi-scale image conversion.

Figure 4 CycleGAN generator structure diagram (see online version for colours)

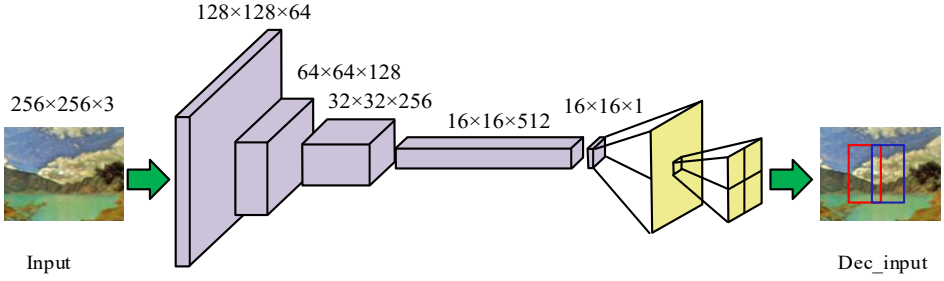


These images are then fed into the adaptive instance normalisation (Pix2Pix) structure for further feature extraction. Pix2Pix technology can transfer the style of one image to another, aligning two encoding layers that need to be fused by merging input features. The main merit of Pix2Pix is its ability to achieve arbitrary style transfer and reduce the number of parameters in the calculation process. The calculation process of Pix2Pix style transfer can be described by equation (9) given the feature vector x of the content image and the feature vector y of the style image.

$$AdalN(x, y) = \sigma(y)\left(\frac{x-\mu(x)}{\sigma(x)}\right) + \mu(y) \tag{9}$$

In equation (9), μ and σ represent the mean and standard deviation calculated for each batch of data. The structure and composition modules of the multi-scale CycleGAN discrimination network are shown in Figure 5.

Figure 5 CycleGAN discriminator structure diagram (see online version for colours)



Corresponding to multi-scale generators, multi-scale discriminators also include multiple discriminators, each responsible for image discrimination of different resolutions. These discriminators work together to ensure accurate differentiation between real images and generated images at different scales. The discriminator part of the multi-scale CycleGAN extracts features from the input image through four convolutional layers and determines the authenticity of the input image through a one-dimensional output convolutional layer.

To make the model easier to deploy on edge computing devices, some of the common convolutional modules in the model are replaced with deeply separable convolutions to reduce the amount of computation of the network parameters, resulting in a lighter model and faster detection. Compared with traditional convolution, depth-separable convolution has fewer parameters. Assuming that the size of the convolution kernel is $h \times w \times c_1$ and has the convolution filling operation, the size of the feature graph $H \times W \times c_1$ after traditional convolution is $H \times W \times c_2$. Depth-separable convolution is a combination of deep convolution and point-by-point convolution. The depth separable convolution is responsible for filtering, its convolution kernel size is $h \times w \times 1$, and the deep convolution has c_1 convolution nuclei acting on each channel. The point-by-point convolution is responsible for transforming the channel, and the size of the convolution kernel is $1 \times 1 \times c_1$. There are c_2 convolution nuclei in the point convolution that act on the output feature map of the depth convolution. Therefore, the parameters of depth convolution $P_{Depthwise}$ and point convolution $P_{Pointwise}$ are calculated as shown in equation (10).

$$\begin{cases} P_{Depthwise} = (h \times w \times 1) \times c_1 \\ P_{Pointwise} = (1 \times 1 \times c_1) \times c_2 \end{cases} \tag{10}$$

The parameter calculation equation of depth-separable convolution $P_{Depthwise\ separable}$ is shown in equation (11).

$$\begin{cases} P_{Depthwise\ separahle} = P_{Depthwise} + P_{Pointwise} = h \times w \times c_1 + c_1 \times c_2 \\ \dfrac{P_{Depthwise\ separable}}{h \times w \times c_1 \times c_2} = \dfrac{1}{c_2} + \dfrac{1}{(h \times w)^2} \end{cases} \tag{11}$$

Depth-separable convolution decomposes traditional convolution, and its parameter calculation is one-tenth of that of traditional convolution. Some common convolutional modules in the network model are replaced by deep separable convolutional modules in the experiment, which significantly reduces the number of model parameters and the model detection delay.

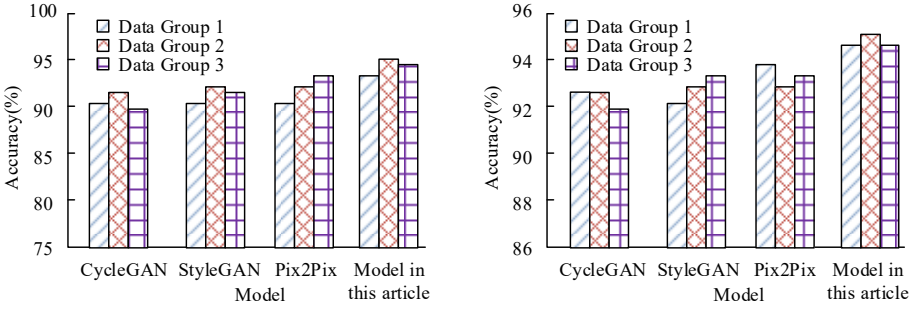
3 Results

3.1 Performance analysis of multi-scale CycleGAN

The experimental environment consisted of an Intel (R) Pentium (R) CPU with a main frequency of 3.60 GHz, 8 GB of memory, 500 GB of hard drive, a Windows 10 operating system host with 110 GB of memory, a 1 TB hard drive, and a 16 GB × 2 NVIDIA Tesla P100 graphics card. The experimental parameters: epoch was set to 200, optimiser was set to ADAM, and initial learning rate was set to 0.001.

The ArtBench dataset was selected for the experiment. This dataset was a class balanced, high-quality, clean annotated, and standardised art generation dataset. The dataset was provided in three versions, each with a different image resolution. The resolution of the images was 32 × 32, 256 × 256, and the original image size. The three versions of the ArtBench-10 dataset were in CIFAR (32 × 32, tar archive), ImageFolder (256 × 256, folder), and LSUN (raw image size, LMDB file) formats, making it easy to use for different machine learning frameworks and image synthesis codebases. The study used Pillow in Python to apply augmentation operations to each image in the training dataset to generate new augmented samples. Even if the content and style images do not match semantically, a large number of training samples can be generated by combining different content and style images. For example, 50 content images can be combined with 50 style images to produce 10,000 different stylised results.

Figure 6 Accuracy comparison, (a) accuracy results of different models on the test set (b) accuracy results of different models on the training set (see online version for colours)



(a) (b)

A comparative analysis was conducted on the performance of multi-scale CycleGAN, traditional CycleGAN, unsupervised image conversion generates adversarial networks (UIC-GAN) and deep convolutional GAN (DCGAN) using evaluation indicators such as accuracy, recall, and F1 score. From Figure 6(a), the accuracy results of the multi-scale CycleGAN on the three versions of the test set were 93.8%, 95.1%, and 94.6%. The proposed model had high accuracy and could effectively identify and classify samples. The results in Figure 6(b) also indicated that multi-scale CycleGAN had the highest accuracy on the training set.

The results in Figure 7 indicated that the proposed algorithm performed well in terms of recall, especially on the test set. Its recall rate on the test set exceeded 90%, with the highest reaching 94.6%. This indicated that the model could effectively identify positive samples, demonstrating the advantage of the proposed algorithm in terms of recall rate compared to these two algorithms.

Figure 7 Comparison of recall rates, (a) recall results of different models on the test set (b) recall results of different models on the training set (see online version for colours)

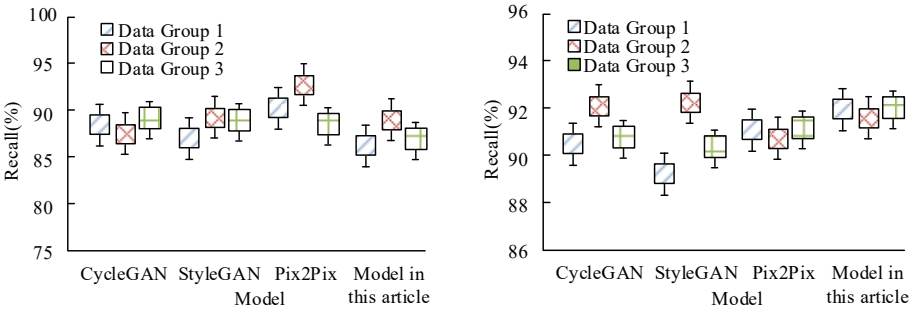


Figure 8 Comparison of F1 values, (a) training set (b) test set (see online version for colours)

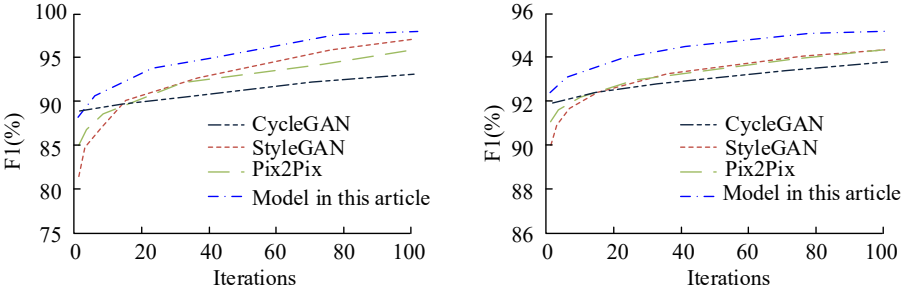


Figure 8 shows that after 100 iterations on the test set, the F1 score of the proposed model converged to 97.6%. The F1 score is a statistical measure that represents the harmonic mean of two key metrics: accuracy and recall. It is a comprehensive indicator that assesses the overall accuracy and recall ability of a given model, making it a valuable metric for evaluating the performance of such models. Compared with the StyleGAN algorithm, the F1 score of the proposed model increased by 2.3%, indicating that on the

test set, the model performed better in balancing accuracy and recall, and could more effectively identify samples while reducing false positives and false negatives.

3.2 Analysis of the application effect of style transfer model in multi-scale CycleGAN

The study validated the efficacy of the style transfer model for multi-scale CycleGAN using three indicators: peak signal-to-noise ratio (PSNR), Frechet inception distance (FID) and inception score (IS). PSNR is a critical indicator for measuring image quality, and the PSNR curves of the three models are shown in Figure 9. From Figure 9(a), the algorithm raised in the training set achieved a PSNR of 95.9 after 80 iterations. UIC-GAN reached the second highest PSNR index value at 85 iterations, 88.7, DCGAN reached the PSNR index value at 90 iterations, and CycleGAN reached the lowest PSNR index value at 90 iterations, 77.6. Figure 9(b) further demonstrated the PSNR performance of the proposed algorithm on the test set, and the PSNR of the raised algorithm was consistently higher than the other two algorithms. This indicated that the proposed algorithm could maintain high image quality on different datasets and had good stability.

Figure 9 PSNR curves of different algorithms, (a) training set (b) test set (see online version for colours)

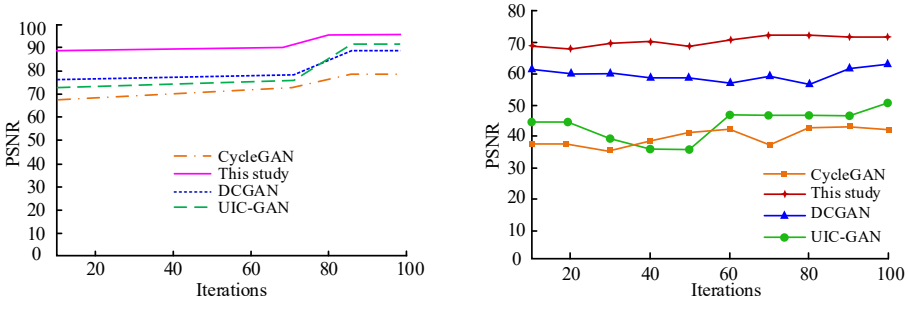
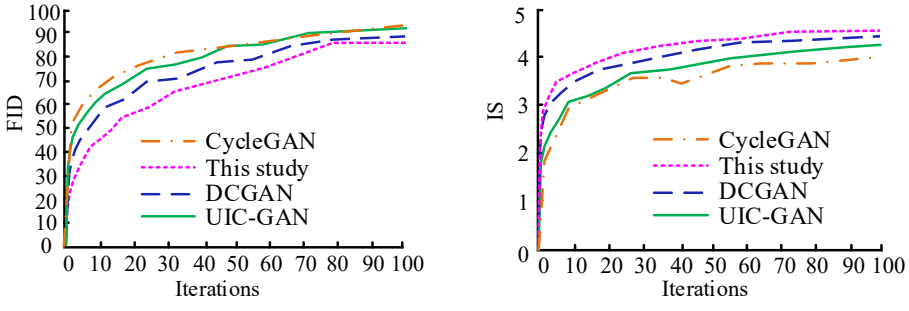


Figure 10 clearly shows that the proposed model consistently maintained a low FID value at different training cycles, indicating that the model had a faster convergence speed compared to the comparative network model. When the model reached the convergence state, through quantitative analysis of objective evaluation indicators, it was found that compared with DCGAN, UIC-GAN and CycleGAN, the FID value of the proposed model decreased by 21.80%, 34.33% and 35.71% respectively, while the IS value increased by 1.755, 0.122 and 1.852 respectively. These results indicated that the network generated illustration images proposed in the study had a statistical distribution that is closer to real image data, exhibiting higher image quality and lower risk of pattern collapse.

Figure 10 Illustration style transfer task FID/IS score comparison (see online version for colours)



(a) (b)

To assess the improvement effect of each module raised in the study, the elimination method was used to investigate the impact of removing a single module on network performance. Here, A1, A2 and A3 represent the loss of low-level texture features, adversarial features, and high-level conceptual features, respectively. B denotes the multi-scale generator, C denotes the multi-scale discriminator, and D denotes the activation function. At the training cycle of 400, ablation experiments were conducted and FID scores were recorded, as shown in Table 1. From the data in the table, the proposed method had the most significant improvement in network performance. When this module was removed, the FID value increased from 73.72 to 102.28, an increase of about 38.74%, which indicated that the fitting ability of the model to the TD was significantly reduced. In addition, removing any of the sub modules A1, A2 and A3 separately would result in an increase in FID values, indicating that these three sub modules that make up the multi-scale adversarial loss can improve model performance. For other modules, when removing the multi-scale generator, multi-scale discriminator, and activation function separately, the FID values increased by 5.90%, 2.63%, and 1.76%, respectively.

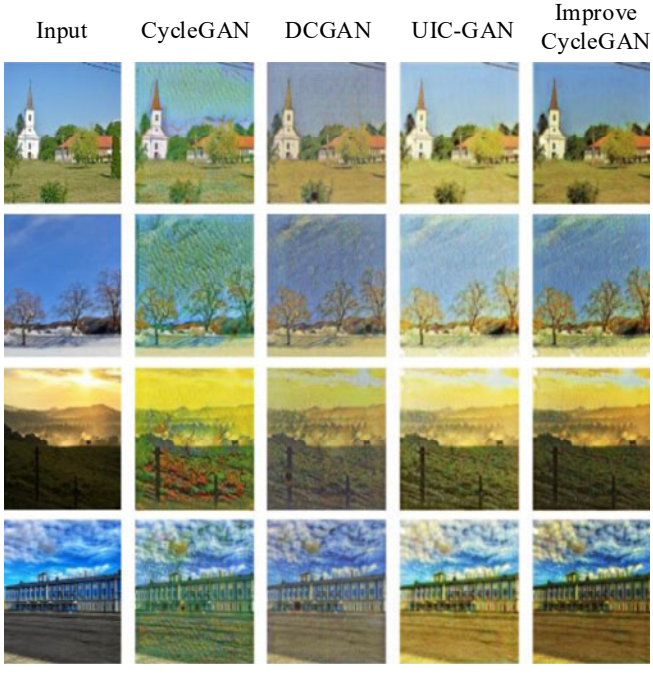
Table 1 Comparison of FID scores in ablation experiments

A1	*A2*	*A3*	*B*	*C*	*D*	*FID*
√	√	√	√	√	√	73.72
	√	√	√	√	√	89.39
√		√	√	√	√	94.23
√	√		√	√	√	79.46
			√	√	√	102.28
√	√	√		√	√	78.07
√	√	√	√		√	75.66
√	√	√	√	√		75.02

Figure 11 shows the contrast map of the transfer effect of illustration image style. In the case of consistent training data and environmental conditions, although traditional cycleGAN can better preserve the content and colour of the source image, insufficient model training may lead to instability and distortion of texture information. The third

column shows that DCGAN enhanced network stability, with no significant distortion in the image, but the colour and structural information was not as good as the basic cycleGAN model. The fourth column indicates that UIC-GAN was not inferior to traditional cycleGAN in learning colour and structural information, but it suffered from texture distortion and colour saturation issues. The fifth column shows the transfer effect of multi-scale cycleGAN illustration style, where the image conversion successfully avoids distortion and displays the best transfer effect.

Figure 11 Animation style transfer contrast diagram (see online version for colours)



4 Conclusions

The study explored the application of multi-scale CycleGAN in style transfer of illustration images. To enhance the effectiveness of style transfer, the existing CycleGAN network structure was optimised. The focus of optimisation was to enhance the performance of the style encoder and discriminator. By improving these two components, the network could more accurately capture and reproduce the target style, while generating more realistic images. Finally, multiple network comparison experiments were conducted on the art style image dataset. The results indicated that the network generated illustration images proposed in the study had a statistical distribution that was closer to real image data, exhibiting higher image quality and lower risk of pattern collapse. Compared with traditional CycleGAN, UIC-GAN and DCGAN, the FID value of multi-scale CycleGAN decreased by 21.80%, 34.33% and 35.71%, respectively, and the IS value increased by 1.755, 0.122 and 1.852, respectively. Multi-scale CycleGAN showed higher stability during training. By introducing multi-scale generator and

discriminator, multi-scale anti-loss function and adaptive style weight mechanism, the model could converge faster and the quality of the generated image was more stable during training. The ablation experiment showed that when removing the low-level texture feature loss, adversarial loss, and high-level conceptual feature loss, the FID value increased from 73.72 to 102.28, an increase of about 38.74%. When removing the multi-scale generator, multi-scale discriminator, and activation function separately, the FID values increased by 5.90%, 2.63% and 1.76%, respectively. Therefore, this model has certain potential for application in the reconstruction of style and content in illustration art expression. Scientificity is an important factor in evaluating the effectiveness of illustration style expression, but this evaluation is limited to qualitative measurement. In the future, a quantitative analysis method for image information preservation can be developed by combining image information measurement, Wasserstein distance, quantitative evaluation factors, and style loss calculation, providing a comprehensive framework to evaluate the results of illustration style transfer.

References

Azni, H.M., Afsharchi, M. and Allahverdi, A. (2023) 'Improving brain tumor segmentation performance using CycleGAN based feature extraction', *Multimedia Tools and Applications*, Vol. 82, No. 12, pp.18039–18058.

Chen, Y.C., Shibata, H., Chen, L.H. and Takama, Y. (2024) 'Synthesis of comic-style portraits using combination of CycleGAN and Pix2Pix', *Journal of Advanced Computational Intelligence and Intelligent Informatics*, Vol. 28, No. 5, pp.1085–1094.

Durrant, J.D. (2022) 'Prot2Prot: a deep learning model for rapid, photorealistic macromolecular visualization', *Journal of Computer-Aided Molecular Design*, Vol. 36, No. 9, pp.677–686.

Gao, Y., Feng, X., Zhang, T., Rigall, E., Zhou, H., Qi, L. et al. (2021) 'Wallpaper texture generation and style transfer based on multi-label semantics', *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 32, No. 3, pp.1552–1563.

Gupta, V., Sadana, R. and Moudgil, S. (2019) 'Image style transfer using convolutional neural networks based on transfer learning', *International Journal of Computational Systems Engineering*, Vol. 5, No. 1, pp.53–60.

Han, H., Yang, B., Zhang, W., Li, D. and Li, H. (2024) 'A modified CycleGAN for multi-organ ultrasound image enhancement via unpaired pre-training', *Journal of Beijing Institute of Technology*, Vol. 33, No. 3, pp.194–203.

Kim, S., Jang, B., Lee, J. and Bae, H. (2023) 'A CNN inference accelerator on FPGA with compression and layer-chaining techniques for style transfer applications', *IEEE Transactions on Circuits and Systems I: Regular Papers*, Vol. 70, No. 4, pp.1591–1604.

Nammee, M. (2023) 'Dog-species classification through CycleGAN and standard data augmentation', *Journal of Information Processing Systems*, Vol. 19, No. 1, pp.67–79.

Richter, S.R., AlHaija, H.A. and Koltun, V. (2022) 'Enhancing photorealism enhancement', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 45, No. 2, pp.1700–1715.

Satchidanandam, A., Al Ansari, R.M.S., Sreenivasulu, A.L., Rao, V.S., Godla, S.R. and Kaur, C. (2023) 'Enhancing style transfer with GANs: perceptual loss and semantic segmentation', *International Journal of Advanced Computer Science & Applications*, Vol. 14, No. 11, pp.321–329.

Sugiyama, S. and Aikawa, N. (2024) 'Detection of painting defects using background subtraction with CycleGAN', *IEEJ Transactions on Electronics, Information and Systems*, Vol. 144, No. 2, pp.80–81.

Wang, J. (2023) 'Garment image style transfer based on deep learning', *Journal of Intelligent & Fuzzy Systems*, Vol. 44, No. 3, pp.3973–3986.

Wang, W., Cui, Z.X., Cheng, G., Cao, C., Xu, X., Liu, Z. et al. (2024) 'A two-stage generative model with CycleGAN and joint diffusion for MRI-based brain tumor detection', *IEEE Journal of Biomedical and Health Informatics*, Vol. 28, No. 6, pp.3534–3544.

Yan, C., Guo, Z. and Cai, Y. (2024) 'Data augmentation of ship wakes in SAR images based on improved CycleGAN', *Journal of Shanghai Jiaotong University (Science)*, Vol. 29, No. 4, pp.702–711.

Yang, S., Kim, E.Y. and Ye, J.C. (2021) 'Continuous conversion of CT kernel using switchable CycleGAN with AdaIN', *IEEE Transactions on Medical Imaging*, Vol. 40, No. 11, pp.3015–3029.

Zhao, S., Chen, X., Yue, X., Lin, C., Xu, P., Krishna, R. et al. (2021) 'Emotional semantics-preserved and feature-aligned cyclegan for visual emotion adaptation', *IEEE Transactions on Cybernetics*, Vol. 52, No. 10, pp.10000–10013.