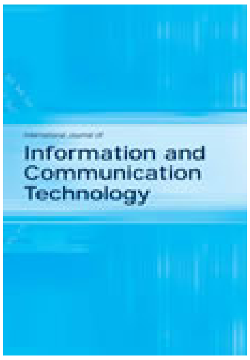


International Journal of Information and Communication Technology

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

Pseudo-coordinates graph convolutional generative adversarial network for art style transfer

Feng Wang, Zihan Zhang

Article History:

Received:	28 November 2024
Last revised:	20 January 2025
Accepted:	20 January 2025
Published online:	31 March 2025

Pseudo-coordinates graph convolutional generative adversarial network for art style transfer

Feng Wang* and Zihan Zhang

Office of Academic Affairs,
Jiangsu Vocational Institute of Architectural Technology,
Xuzhou, Jiangsu, 221116, China
Email: wangfeng20242024@163.com
Email: zhangzihan9@jsjzi.edu.cn
*Corresponding author

Abstract: The goal of style transfer is to apply the artistic features of a style image to a content image while maintaining the content image's structure. Traditional methods often use CNNs and residual blocks, but their limited receptive field struggles to capture long-range feature dependencies, leading to repetitive local patterns. Residual blocks can also cause interference between style and content representations. To address these issues, we introduce a pseudo-coordinates graph convolutional generative adversarial network (PGC-GAN), which consists of two branches: one for extracting style and another for style transfer. The style extraction branch represents style features as a graph and uses graph pooling to remove redundant information. The style transfer branch encodes these features into pseudo-coordinates, enabling flexible relationships between pixel nodes and long-range feature aggregation without disrupting the content image's structure. Experimental results demonstrate that PGC-GAN significantly improves artistic style transfer compared to existing methods.

Keywords: art style transfer; graph convolution; pseudo-coordinates; GAN.

Reference to this paper should be made as follows: Wang, F. and Zhang, Z. (2025) 'Pseudo-coordinates graph convolutional generative adversarial network for art style transfer', *Int. J. Information and Communication Technology*, Vol. 26, No. 6, pp.45–61.

Biographical notes: Feng Wang is an Associate Professor and has a Master's degree. She graduated from Soochow University in 2001 and worked in Jiangsu Vocational Institute of Architectural Technology. Her research interests include design science.

Zihan Zhang is an Assistant Researcher and has a Master's degree. He graduated from Chongqing University in 2020 and worked in Jiangsu Vocational Institute of Architectural Technology. His research interests include image.

1 Introduction

Artistic style transfer is a crucial task in computer vision, with wide applications in image editing and digital art creation. The core objective is to transfer the artistic style of a target image onto a source image while preserving the content of the original image, resulting in an image that embodies the desired style (Wang and Yaermaimaiti, 2024). However, challenges arise from the differences between styles, the degree of style-content fusion, and the need to maintain image details during the transfer process (Wei, 2024; Tian, 2024; Liu, 2023). These challenges limit the quality and consistency of images generated by current style transfer methods.

Artistic style transfer methods based on CNNs (Ni, 2024; Wang and Li, 2023; Farajzadeh et al., 2023; Chiu and Gurari, 2023) rely on their hierarchical convolutional structure, which can automatically extract multi-scale features from images, offering advantages in capturing local details while preserving global structures. Li and Zhu (2024) proposed a multi-scale feature fusion network that employs a parallel multi-scale feature extraction mechanism to capture style features at different levels. By incorporating channel attention mechanisms, they emphasise the significant features of the target artistic style. However, to reduce the number of model parameters, Li and Zhu (2024) used a convolutional kernel decomposition strategy, breaking down large kernels into smaller ones to expand the receptive field and lower computational complexity. Although stacking smaller kernels can increase the receptive field to some extent, it also raises training challenges, such as gradient explosion or vanishing. To address this issue, Chiu and Gurari (2022) introduced high-frequency residual skip connections, transmitting high-frequency details between the encoder and decoder. This method preserves more details through residual structures, ensuring that high-frequency features of the original image are retained during style transfer. However, despite preserving high-frequency information, these details do not entirely equate to style features. Since edges and other high-frequency information are closely tied to image content, models struggle to distinguish between these details and style information, leading to confusion between content and style in the transfer results.

Compared to CNN-based methods, generative adversarial networks (GANs) (Lin and Li, 2024) offer a different approach to style transfer. GANs, through adversarial training between the generator and discriminator, effectively mitigate gradient explosion or vanishing issues. Additionally, the multi-stage generation structure of GANs allows images to be generated progressively, which aids in separating content and style features while maintaining global consistency. For example, Han et al. (2023) proposed a multi-feature encoder that encodes style from multiple aspects, including shape, texture, and colour, and utilises dynamic convolution and adaptive instance normalisation (AdaIN) for effective transfer of complex styles. Ma (2024) further explored different generator architectures in GANs and discovered significant differences in style transfer performance depending on the architecture. For instance, using U-Net (Ronneberger et al., 2015) as a generator, the model excels in retaining details and textures due to its complex structure. However, U-Net tends to generate overly detailed and textured images, resulting in overly complex and visually cluttered outputs. On the other hand, using ResNet (He et al., 2016) as a generator produces more natural and consistent style representations but lacks the detailed textures and complexity often crucial to artistic style due to ResNet's residual structure, which directly preserves some of the original content information.

To further optimise style transfer performance, some methods have introduced graph structures to more flexibly represent relationships between style features. This approach aligns well with the characteristics of style features, which are often independent of the absolute position in the image. Shi et al. (2024) proposed a heterogeneous graph structure that includes both the target image and multiple reference images. Through the similarity reference indexing generation module, it selects reference images that are semantically and stylistically similar to the target image. The multi-reference graph reasoning module then uses GCNs to learn the relationships between images within the graph, optimising the style transfer process. Similarly, Jing et al. (2022) utilised graph neural networks (GNNs) to establish fine-grained content-style correspondences, treating local patches from both content and style images as graph nodes. They employed a heterogeneous graph attention mechanism to enable patch-level content-style interactions, thereby forming adaptive many-to-one content-style correlations. Additionally, a deformable graph convolution method was introduced to achieve cross-scale content-style matching.

However, defining relationships between nodes in graph structures remains a challenge. Node relationships are often based on similarity or a fully connected approach. The former has limited flexibility and tends to fall into local optima (Jiao et al., 2022), while the latter, although capable of capturing global semantic correlations, leads to node oversmoothing as the number of layers increases, resulting in feature information loss (Cao et al., 2022). Although attention-based node relationships improve flexibility to some extent, the use of Softmax in attention mechanisms forces weights to be non-negative and sum to 1, which can lead to over-squashing during information aggregation, causing the loss of critical local information (Giraldo et al., 2023).

Therefore, we believe that a key challenge in improving the quality of generated images is how to effectively separate and integrate style and content features while ensuring consistent and uniform style distribution across the image.

To overcome the limitations of existing style transfer methods in separating and integrating style and content features, we propose an innovative network architecture called the Pseudo-coordinates graph convolutional generative adversarial network (PGC-GAN). This architecture effectively addresses the issues of poor style consistency and redundant information in traditional methods by separating style extraction and style transfer into two independent branches. In the style extraction branch, we model style features using a graph structure, enabling the relationships between nodes to dynamically adapt to different styles and enhancing style consistency. To tackle the difficulty of eliminating redundant style features in traditional methods, we designed a node scoring mechanism (NSM) that assigns a uniqueness score to each node. Nodes with higher scores represent more distinctive style information, while nodes with lower scores are considered redundant and can be represented by combinations of other nodes. This mechanism effectively reduces invalid or repetitive style features, preventing interference during the style transfer process. In the style transfer branch, we abandon the traditional residual structure to overcome the issue of mixed style and content features in previous methods. However, removing the residual structure may lead to gradient vanishing or explosion, and reducing convolutional blocks can shrink the receptive field, weakening the model's ability to capture global features. To address these challenges, we replace convolutional blocks with graph convolutional networks (GCNs) (Zhang et al., 2022) to enhance the model's capacity for modelling complex style features. However, traditional GCNs have limitations in constructing node relationships: similarity-based connections are too simplistic to capture complex global relations, while fully connected graphs, though capable of capturing global information, increase computational complexity and lead to node over-smoothing. To solve this problem, we propose a learnable pseudo-coordinates graph convolution (PGC) mechanism. This method dynamically learns flexible relationships between nodes based on style feature encoding, avoiding reliance on the original structure of the content image and effectively balancing local and global information modelling. By precisely modelling long-range dependencies, PGC not only improves the accuracy and consistency of style transfer but also reduces computational complexity, overcoming the shortcomings of traditional methods in node relationship modelling.

In summary, our contributions are threefold:

1. We propose a novel pseudo-coordinates PGC-GAN that aggregates long-range style dependencies while preserving the original structure of the content image.
2. We design a NSM for the style extraction branch, which scores each graph node to identify and filter out redundant style feature nodes, thereby reducing redundant style information and ensuring consistency in style.
3. We introduce PGC for generating pseudo-coordinates for pixel nodes based on style features. This further separates dependencies between the original image and style image, allowing precise control over long-range dependencies during the style transfer process.

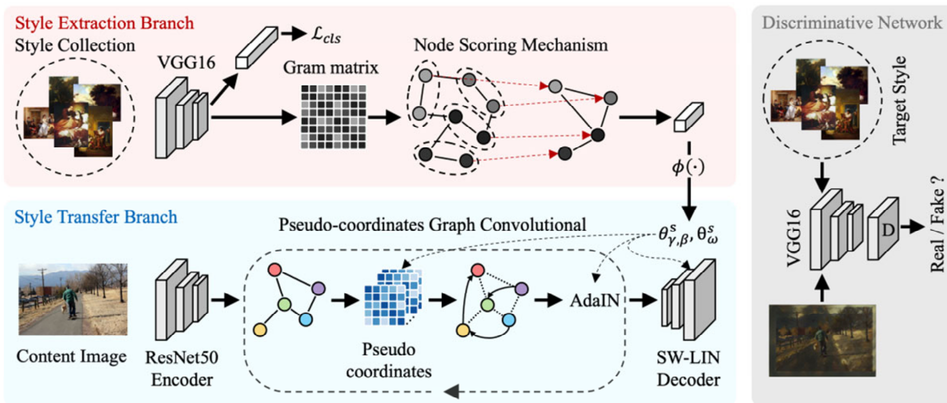
2 Methods

In this section, we will introduce the proposed method in detail. First, we describe the overall framework of the model, as shown in Figure 1. Then, we provide a more detailed explanation of the Pseudo-coordinates PGC-GAN, focusing on the key components of the two branches: the NSM and the PGC.

The PGC-GAN model consists of three parts, as illustrated in Figure 1. First, the style extraction branch extracts the style information from the style collection. Then, the style transfer branch extracts the content features from the content image, fuses them with the extracted target style, and generates the style-transferred image. Finally, the discriminator network judges whether the generated image's style is consistent with the target style.

Specifically, we utilise the VGG16 and ResNet50 networks pre-trained on ImageNet (Deng et al., 2009) as feature extractors for the style extraction and style transfer branches, respectively. First, we choose VGG16 as the feature extractor for the style extraction branch due to its relatively simple structure, which lacks skip connections, thus avoiding confusion between content and texture features. In this branch, we extract features using VGG16 and calculate the Gram matrix to represent style features. Each style is treated as a node, and we calculate the importance score for each node, with higher scores indicating that the node contains more information, while lower-scored nodes can be replaced by others. This process filters out redundant style features. Ultimately, the style features are modelled as two sets of parameters that guide the style transfer branch.

Figure 1 Framework of the pseudo-coordinates graph convolutional generative adversarial network (see online version for colours)



In the style transfer branch, ResNet50 is responsible for extracting content features. Its residual blocks can effectively reuse lower-level features, aiding in content extraction. We treat each pixel in the feature map extracted by ResNet50 as a node and use the topology parameters generated from the style features to construct the graph structure between nodes. By introducing pseudo-coordinates, we enhance the flexibility of graph construction, while avoiding excessive node smoothing. Next, we aggregate features based on the pseudo-coordinates and perform L rounds of aggregation, followed by AdaIN (Huang and Belongie, 2017). It is important to note that the parameters for AdaIN are also generated from the style features. Finally, the style-transferred image is generated through a decoder equipped with a spatial window layer instance normalisation (SW-LIN) (Xu et al., 2021) function.

In the discriminator network, VGG16 takes both the target style image and the generated image as inputs. The discriminator is used to determine whether the generated image's style is consistent with the target style.

2.1 Node scoring mechanism

To avoid degradation in the quality of generated images due to redundant local style features, we designed and proposed a NSM. This mechanism aims to evaluate the importance of each node by calculating the Manhattan distance between nodes, determining which nodes' information can be represented or replaced by others, thereby reducing redundant features. This approach not only enhances the model's ability to handle global information but also optimises the style transfer process, ensuring that the generated images exhibit a more harmonious and unified overall style.

First, we base our approach on the feature map $F \in R^{H \times W \times C}$ extracted from the VGG16 network, where this feature map contains high-dimensional content representations. To process these high-dimensional features, we calculate the Gram matrix $G \in R^{C \times C}$ to capture the relationships between style features. The calculation of the Gram matrix is expressed in equation (1):

$$G(i, j) = \sum_{k=1}^{H \times W} F_i k F_j k \quad (i, j = 1, 2, \cdots C) \tag{1}$$

where F_{ik} represents the k^{th} element of the i^{th} channel in the feature map. Next, we treat each row of the Gram matrix as a feature vector node and represent these nodes as $v_c \in R^{1\times C}$. By calculating the similarity between nodes, we construct the adjacency matrix $A \in R^{C\times C}$ and generate the corresponding degree matrix $D \in R^{C\times C}$. To further filter out redundant nodes, we introduce the Manhattan distance as a metric to quantify the importance of each node. Specifically, the score $p \in R^{1\times C}$ for each node is calculated using the following equation (2):

$$p = \left\| \left(I - D^{-1} A \right) V \right\|_1 \tag{2}$$

where I is the identity matrix, V represents the node feature matrix, and $\|\cdot\|_1$ is the L1 norm operator applied row-wise. After computing the scores for each node, we sort these scores and introduce a selection ratio r to determine the number of nodes to retain. The specific node selection process is described by equation (3):

$$dx = top - rank(p, r * C), \tilde{V} = V(idx, :), \tag{3}$$

where *top-rank*$(\cdot)$ is a function that returns the indices of the top $\lceil r*C \rceil$ values. The selected features $\tilde{V} \in R^{[r*C]\times C}$ are then encoded to guide the subsequent style transfer module, with the parameter values computed as shown in equation (4):

$$\left\{ \theta_\omega, \theta_{\gamma,\beta} \right\} = \left\{ BN\left(j_\omega \left(pool(\tilde{V}) \right) \right), BN\left(j_{\gamma,\beta} \left(pool(\tilde{V}) \right) \right) \right\} \tag{4}$$

where ϕ_ω and $\phi_{\gamma,\beta}$ are fully connected layers used for encoding, ϕ_ω constructs the pseudo-coordinates, and $\phi_{\gamma,\beta}$ generates the affine parameters for the AdaIN layer. BN denotes batch normalisation. It is important to note that since we use a style collection where each style contains K style images, the parameters generated by the encoding process are weighted averages. This process is described by equation (5):

$$\left\{ \overline{\theta}_\omega, \overline{\theta}_{\gamma,\beta} \right\} = \left\{ \frac{1}{K} \sum_{k=0}^{K} \pi_k \theta\omega_k, \frac{1}{K} \sum_{k=0}^{K} \pi_k \theta_{\gamma k}, _{\beta k} \right\} \tag{5}$$

where π_k denotes the weight corresponding to the k^{th} style image, which is determined by the similarity between the style image and the content image.

2.2 Pseudo-coordinates graph convolutional

In style transfer, traditional CNNs face issues such as a large number of parameters, complex training, and feature confusion. In contrast, GCNs can effectively facilitate feature aggregation between nodes, helping to share style information across different nodes. However, due to the complexity of constructing graph structures, GCNs have higher computational costs and lower interpretability. To address these issues, we introduce pseudo-coordinates based on conventional graph convolution.

Specifically, we first reconstruct the parameter $\overline{\theta}_\omega$, obtaining the adjacency matrix $A \in R^{HW\times HW}$, where H and W represent the height and width of the feature map output by ResNet50. Next, we compute the degree matrix $D_{i,i} = \sum A_{i,j}$ and the random walk matrix $M = D^{-1}A$, where $M_{i,j}$ represents the probability of moving from node i to node j in one random walk. Based on this concept, we construct pseudo-coordinates $P_{i,j}$, with the calculation process described by equation (6):

$$P_{i,j} = \left[I, M, M^2, \cdots, M^{I,M,M^2,?,M^{(K-1)}} \right]_{i,j} \in R^K, \tag{6}$$

where $P_{i,j}$ represents the probability of node i reaching node j in K-1 steps. This construction of pseudo-coordinates captures the unique relationship between nodes i and j and is better at capturing multi-step dependencies compared to other methods.

Next, since using the same weights for each channel in multiple graph convolutions can lead to over-smoothing, we introduce a kernel function $\psi(\cdot)\in R^{K\times d}$ to ensure that each channel has different weights. For the graph $G = (V, E)$ and its node signal function: $V \rightarrow R^d$, the PGC is defined as follows in equation (7):

$$(\chi * \psi)(i) := W\left(\frac{1}{\left|supp_\psi(i)\right|} \sum\nolimits_{j\in supp(i)} \chi(j) \odot \psi\left(P_{i,j}\right) \right) + b \tag{7}$$

where b is a bias term, W is the trainable weight, and $supp_\psi(i)$ represents the K-hop neighbourhood of node i, i.e., the set of nodes reachable from i with a probability greater than or equal to ε after K steps.

The aggregated node features X are then normalised using AdaIN to match the mean and variance of the style input $\overline{\theta}_{\gamma,\beta}$, with the calculation process described by equation (8):

$$AdaIN\left(X, \overline{\theta}_{\gamma,\beta}\right) = \overline{\theta}_\gamma \left(\frac{X - \mu(X)}{\sigma(X)} \right) + \overline{\theta}_\beta, \tag{8}$$

where $\overline{\theta}_\beta$ and $\overline{\theta}_\gamma$ represent the mean and variance of the style features. After performing L graph convolutions, the output features X are fed into the SW-LIN decoder for upsampling, ultimately generating the style-transferred image I_s. The SW-LIN decoder is a symmetric structure of ResNet50, removing skip connections and replacing BN layers with AdaIN layers.

Finally, we optimise the discriminator with respect to the target style image and extract features from both the generated image and the target style image to evaluate their style consistency in feature space.

2.3 Loss function

During training, we use adversarial loss L_{adv}, perceptual loss L_{per}, and style classification loss L_{cls}. The overall loss function calculation is given by equation (9):

$$L = L_{adv} + \lambda_{per} L_{per} + \lambda_{cls} L_{cls} \tag{9}$$

where λ_{per} and λ_{cls} are weight parameters.

First, adversarial loss aims to assess whether the generated image is similar to images in the target style collection, with the calculation process described by equation (10):

$$L_{adv} = E_{y^c, y_i^c \sim Y, c\sim N}\left[-logD\left(y^c, \left\{y_i^c\right\}_{i=0}^M\right) + E_{\tilde{x}^c \sim G(x), y_i^c\ Y, c\sim N} \right]\left[-log\left(1 - D\tilde{x}, \left\{y_i^c\right\}_{i=0}^M\right)\right] \quad 10)$$

where c represents the c^{th} style out of N different styles, y denotes real images, and x denotes generated images. M represents the number of images in the target style collection, which we set to 3 in our experiments.

Next, perceptual loss aims to compute style loss L_s and content loss L_c between the generated image and content image from multiple levels, as shown in equation (11):

$$L_{per} = \lambda_c L_c + \lambda_s L_s \tag*{11)}$$

where the content loss and style loss are calculated as shown in equations (12) and (13):

$$\begin{aligned} L_c &= E_{x\sim X, c\sim N} \left\| \varphi(x) - \varphi\left(\tilde{x}^c\right) \right\|, \\ L_s &= E_{l\, N_l} \left(\left(\mu^l_{y^c} - \mu^l_{\tilde{x}^c} \right)^2 + \left(\left(Gram^l_{y^c} - Gram^{l\;2}_{\tilde{x}^c} \right)^2 \right) \right) \end{aligned} \tag*{13)}$$

where $\varphi(\cdot)$ denotes features extracted from the $ReLU_{41}$ layer of the VGG network, and L_s is applied to the l^{th} layer features of VGG, which are $ReLU_{12}$, $ReLU_{22}$, $ReLU_{33}$, $ReLU_{43}$, and $ReLU_{51}$ layers respectively.

Finally, style classification loss aims to assist the style extraction network in classifying styles to extract more accurate style features, with the calculation process described by equation (14):

$$L_{cls} = E_{y\sim Y, c\sim N} \left(-logD_{cls} \left(c \mid y^c \right) \right) \tag{14}$$

3 Results and discussion

3.1 Datasets and evaluation metrics

We use the Place365 dataset (Zhou et al., 2017) and the WikiArt dataset (Mohammad and Kiritchenko, 2018) as sources for content images and style images, respectively. Specifically, 56,287 images were randomly selected from Place365, covering 69 different scenes. The WikiArt dataset contains 107,729 artworks across 15 different painting genres. To evaluate the effectiveness of style transfer, we use the deception score (DS) as the evaluation metric. The DS measures the percentage of correct predictions by a pre-trained artist classification network when the stylised images are input, assessing the quality of style transfer in the generated images. Additionally, in ablation studies, we introduce style distance (SD) to further evaluate the effectiveness of each component. The SD is calculated similarly to equation (13), using the L2 norm of the Gram matrices and summing them.

3.2 Implementation details

Our method is implemented using the PyTorch deep learning framework and trained on an NVIDIA RTX 4090D GPU. During training, images are augmented with random rotations and flips and then cropped to a resolution of 768 × 768. For the NSM module, the selection ratio r is set to 0.9. In the PGC module, the number of iterations L is set to 16, the K-hop neighbourhood range covers the entire graph, and the probability threshold ε is set to 0.7. We use the Adam (Diederik, 2014) optimiser with a learning rate of 0.0001, a batch size of 2, and the model is trained for a total of 360,000 iterations.

3.3 Ablation study

To validate the effectiveness of the proposed NSM and PGC methods, we tested these components on selected image pairs, and the results are shown in Table 1.

Table 1 Performance comparison of the proposed PGC-GAN, and the impact of NSM and PGC

Method	*NSM*	*PGC*	*SD↓*	*DS↑*
1	×	×	284.5	0.450
2	√	×	271.8	0.536
3	×	√	263.4	0.574
PGC-GAN	√	√	236.1	0.612

fThe results in Table 1 indicate that both NSM and PGC significantly improve the model's performance. In Method 1, the style extraction branch uses features extracted by VGG16 directly encoded for the style transfer branch's AdaIN, while the style transfer branch replaces PGC with CNN residual blocks with skip connections. In this case, the model shows high SD and low DS, reflecting mediocre performance. After introducing NSM, SD decreases to 271.8, and DS increases to 0.536. This demonstrates that NSM effectively reduces redundant style features, improving the model's style consistency and transfer effect. A decrease in SD indicates reduced style discrepancy and more unified style, while an increase in DS suggests that the generated image better matches the target style. Following the introduction of PGC, SD further drops to 263.4 and DS rises to 0.574. PGC, by constructing pseudo-coordinates and graph convolutions, allows the model to better capture long-range feature dependencies, enhancing the precision and accuracy of style transfer and further improving the style features of the generated images. Finally, with the combined introduction of NSM and PGC, SD falls to 236.1, and DS rises to 0.612, achieving the best results. NSM primarily targets the style extraction stage by assigning a uniqueness score to each node, effectively filtering out redundant and irrelevant style features. Traditional style transfer methods often include a large amount of repetitive or unrelated information during style extraction, leading to unstable or even distorted transfer results. By preserving representative style features, NSM reduces interference between style features, enabling the model to generate images with more coherent and consistent style representation. On the other hand, PGC further addresses the limitations of traditional models in modelling global and long-range feature dependencies during the style transfer process. Conventional CNN structures and GCNs struggle with capturing long-range dependencies, often resulting in localised style information that compromises overall style consistency. PGC overcomes this limitation by constructing pseudo-coordinate graphs and employing flexible graph convolution mechanisms, enhancing the model's ability to perceive and integrate cross-regional style features. Overall, the designs of NSM and PGC improve model performance from two critical aspects: eliminating redundant features and enhancing global information modelling. Their synergistic effect significantly boosts the effectiveness of style transfer.

Additionally, to further analyse the impact of NSM and PGC on model performance, we compared stylised results for selected images, as shown in Figure 2. The images are as follows:

a the original content image

b the reference style image

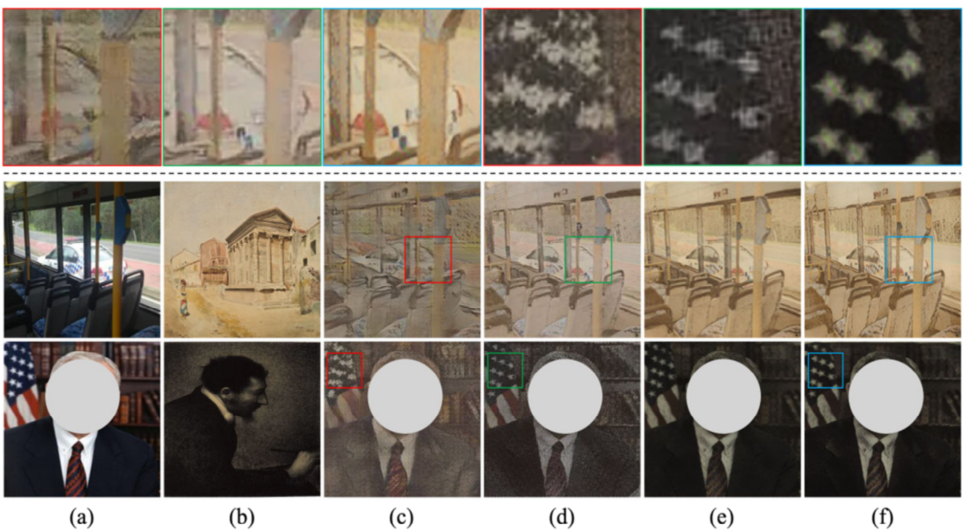
c results from Method 1

d results from Method 2

e results from Method 3

f results from the PGC-GAN model.

Figure 2 Ablation study of PGC-GAN (see online version for colours)

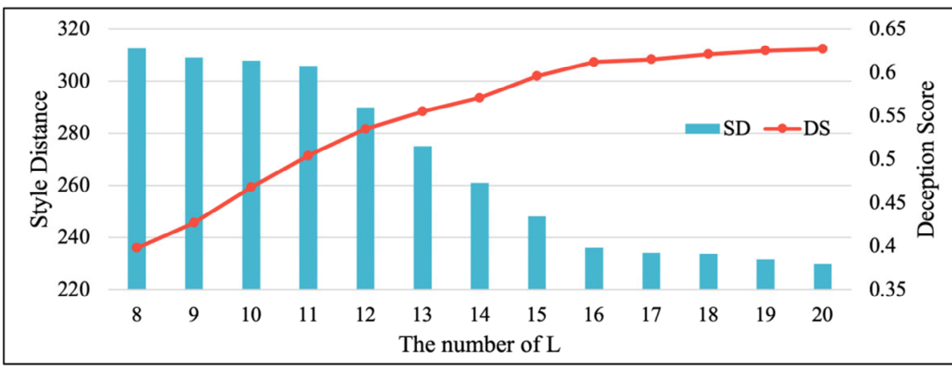


In the visual analysis of the ablation experiments shown in Figure 2, different methods exhibit significant differences in the visual quality of the stylised images. In Method (c), without the introduction of NSM and PGC, the generated images display obvious content confusion, blurred details, and distorted boundaries. This is due to the lack of NSM for filtering redundant style features, resulting in interference between style and content features, and the absence of PGC for long-range feature aggregation, preventing the model from effectively capturing global information and further weakening the representation of image details and structure. In Method (d), with only NSM introduced, the colour style consistency of the images is noticeably improved. NSM effectively reduces redundant features and minimises interference between style and content. However, due to the lack of PGC's ability to model long-range features, the images still suffer from blurred details and unclear boundaries. In Method (e), with only PGC introduced, the images show significant improvements in detail representation and structural boundaries compared to Method (c). PGC enhances long-range feature aggregation through the construction of pseudo-coordinate graphs and graph convolution mechanisms, resulting in clearer textures and edges. However, without NSM to filter redundant features, local style inconsistencies still exist. In contrast, Method (f), which incorporates both NSM and PGC, significantly improves the overall image quality. NSM effectively extracts key style features and prevents interference from redundant information, while PGC enhances detail and structure representation through long-range feature aggregation. The combination of both not only makes the images more naturally unified in colour and texture but also greatly improves detail clarity and style layering, achieving the best style transfer results.

3.4 Quantitative analysis

To further investigate the impact of hyperparameters on model performance, we tested different values for the selection ratio *r*, the number of iterations *L* in the PGC module, and the probability threshold *ε*. The number of iterations *L* in the PGC module directly affects the degree of content and style fusion, which significantly impacts the final generated image. Therefore, we conducted experiments with various values of *L*, and the results are shown in Figure 3.

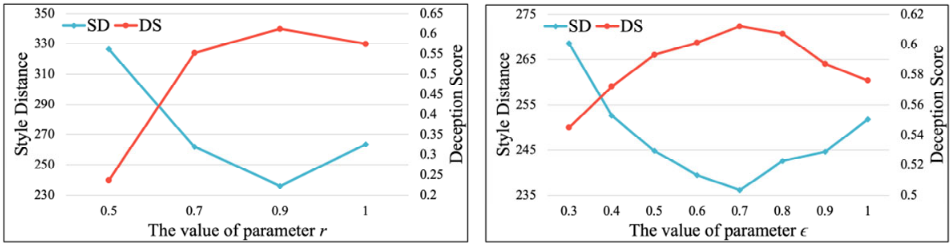
Figure 3 Impact of the number of iterations *L* in the PGC module on model performance (see online version for colours)



The number of iterations *L* in the PGC module plays a crucial role in style transfer, directly affecting the fusion of content and style, and thus determining the quality of the generated image. When *L* is less than 16, feature aggregation is insufficient, leading to incomplete fusion of style and content, and the generated images lack complex details and clear hierarchical structures. As *L* increases, the model gradually captures more long-range feature dependencies, and the image quality improves significantly. However, when *L* exceeds 16, performance gains begin to plateau, and further increases in continue to improve results but at the cost of computational efficiency and model complexity. Therefore, to balance performance and computational cost, *L* is set to 16.

The selection ratio *r* and probability threshold ε determine the proportion of retained style feature nodes in NSM and the number of neighbouring nodes in PGC, respectively, impacting the style consistency and content detail of the generated image. We tested these two hyperparameters, and the results are shown in Figure 4.

Figure 4 Impact of the selection ratio r and probability threshold ε on model performance (see online version for colours)



When the selection ratio r is too small, the model retains too few style feature nodes, which cannot adequately express the diversity of the target style, leading to incomplete style representation in the generated images, with a lack of rich details and layers. This also weakens the style transfer effect, causing the style features in the generated images to be less distinct and the fusion of style and content to be poor. For the probability threshold ε, if ε is too large, the model only considers nearby nodes, failing to capture relationships between distant features effectively. This limits the global style consistency of the generated images, resulting in better local style representation but lacking overall style coherence, with insufficient global information fusion and affected detail processing. When ε is too small, the model considers too many distant nodes, which enhances global feature aggregation but may lead to excessive smoothing of features, reducing image detail representation and causing confusion between style and content features, affecting image clarity and style transfer effectiveness. Additionally, a small threshold increases model complexity, leading to decreased training and inference efficiency. Therefore, after experimental testing, we set the selection ratio r and probability threshold ε to 0.9 and 0.7, respectively.

We then compared the proposed PGC-GAN with existing state-of-the-art (SOTA) methods, and Table 2 presents the comparison results.

Table 2 Performance comparison of the proposed PGC-GAN with other SOTA methods

Method	*SD↓*	*DS↑*	*Parameters (M)*	*FLOPs (G)*
Clipstyler (Kwon and Ye, 2022)	271.8	0.561	48.5	62.3
AesUST (Wang et al., 2022)	258.2	0.568	54.7	71.8
IEContraAST (Chen et al., 2021a)	263.4	0.552	43.2	58.6
Dualast (Chen et al., 2021b)	268.3	0.542	49.8	66.4
SAE-CGM (Xu et al. 2023)	245.6	0.573	52.1	69.2
PGC-GAN	236.1	0.612	55.4	73.5

Our proposed PGC-GAN achieves the best performance, while the closest competitor, the SAE-CGM model, implements the ‘style kernel’ mechanism to facilitate global and local feature interaction, ensuring flexibility in style transfer and preservation of content structure. However, SAE-CGM has limitations in balancing global and local features, particularly when overly focusing on local details, which can lead to uneven style feature distribution and style leakage issues. Additionally, SAE-CGM primarily relies on global style-content alignment features to generate the ‘style kernel,’ which lacks flexibility and accuracy in handling complex long-range dependency features. In contrast, our

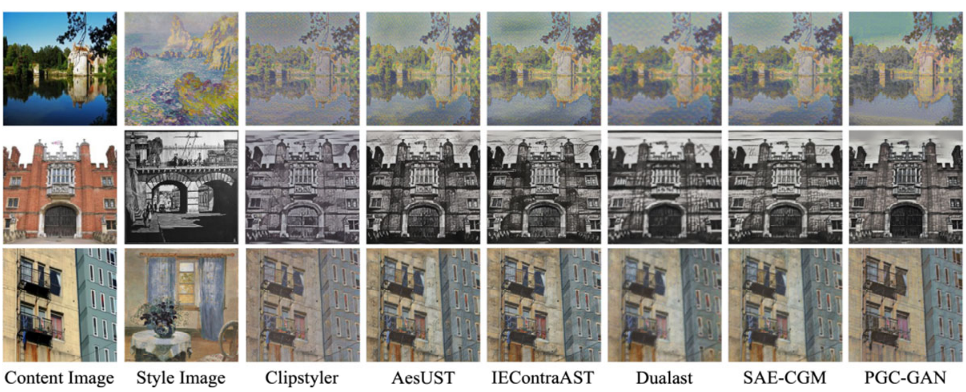
PGC-GAN, by introducing PGC, better captures long-range dependency features, enhances the fusion of style and content, and avoids excessive local feature dominance in image style, achieving better global and local balance. Moreover, the NSM in PGC-GAN further optimises the selection of style features, eliminating redundant information and making style transfer more precise.

In terms of model complexity, PGC-GAN maintains a competitive balance between performance and computational efficiency. With 55.4 million parameters and 73.5 GFLOPs, PGC-GAN achieves superior results while keeping the model size and computational demand within a reasonable range. In comparison, SAE-CGM has 52.1 million parameters and 69.2 GFLOPs, reflecting slightly lower complexity but at the cost of reduced flexibility in feature aggregation and style transfer accuracy. Despite PGC-GAN having marginally more parameters and computational load, its design effectively utilises these resources to achieve better global-local feature integration and overall style transfer performance. These improvements enable PGC-GAN to exhibit higher flexibility and efficiency in style transfer without significantly increasing computational overhead.

3.5 Qualitative analysis

Finally, we visually compared the style-transferred images generated by various SOTA models to assess the quality of the generated images. The comparative results are shown in Figure 5.

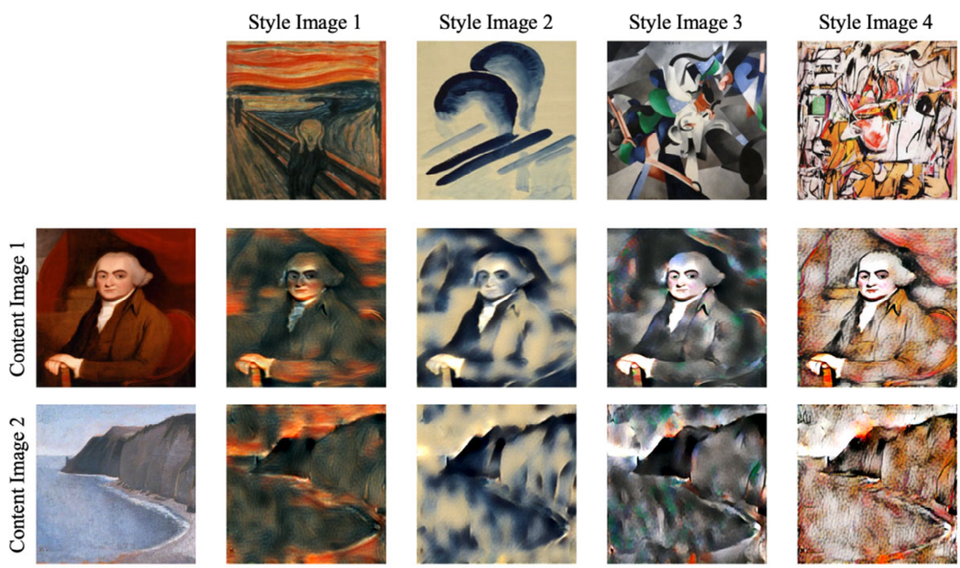
Figure 5 Comparison of style-transferred images generated by the proposed PGC-GAN and other SOTA methods (see online version for colours)



In the comparison of generated images, noticeable artefacts are present in the results from the Clipstyler and IEContraAST models, such as black streaks in the sky. Clipstyler's issue arises from its heavy reliance on text descriptions for style transfer, leading to insufficient handling of local details. This is particularly problematic when no reference style image is available, as the CLIP-based text-image matching loss results in unstable texture generation. IEContraAST, which uses internal and external contrastive losses to enhance style consistency, also falls short in local detail processing, resulting in unrealistic colour and texture distributions. AesUST, while producing results close to the original style image, focuses excessively on texture, particularly in its aesthetic enhancement module AesSA, which overemphasises textures in large areas such as the sky and walls, neglecting smooth colour transitions and causing colour distortion. DualAST struggles with an imbalance in its dual-style learning framework for global and local style control, leading to image blurring and confusion between style and content, especially during complex style transitions. Although SAE-CGM employs a 'style kernel' mechanism to dynamically generate convolutional kernels, it lacks precision in detail texture generation, resulting in detail loss and impacting the final output. In contrast, our proposed PGC-GAN effectively captures long-range feature dependencies and balances global-local features through PGC and optimises style feature selection with NSM. It significantly outperforms the aforementioned methods in terms of colour and texture representation.

To further evaluate the robustness and generalisation ability of the proposed PGC-GAN, we conducted experiments on images with various styles and complexities. These tests assessed the model's capability to adapt to different stylistic features while preserving content structure and achieving consistent style transfer. The experimental results are shown in Figure 6.

Figure 6 The generalisation performance of the proposed PGC-GAN on images of different styles (see online version for colours)



The experimental results in Figure 6 demonstrate that the proposed PGC-GAN exhibits strong adaptability to various artistic styles, including the dramatic colours and expressive textures in Style Image 1, the soft tones in Style Image 2, the geometric abstract features in Style Image 3, and the complex layered textures in Style Image 4. At the same time, the core structure of the content images, such as portraits or landscapes, is well preserved despite significant style variations. The model achieves long-range feature aggregation through the PGC mechanism, effectively balancing global style consistency with the preservation of local details. Additionally, the NSM eliminates redundant features, making style transfer more coherent and precise. However, in certain cases, such as the geometric style transfer in Style Image 3, slight fusion artefacts appear in the landscape content, while in Style Image 2, some detailed textures appear slightly smoothed. Overall, PGC-GAN performs remarkably well across various styles, validating its robustness and generalisation capabilities while maintaining consistency in content structure and style representation.

4 Conclusions

To address the challenges in painting style transfer tasks – such as the integration of style and content, capturing long-range feature dependencies, and handling redundant style features-we propose an innovative Pseudo-coordinates PGC-GAN. This model effectively aggregates long-range style features while preserving the original structure of content images. Additionally, we designed the NSM to optimise the selection of style features and employed PGC to learn flexible node relationships. Extensive experimental results demonstrate that PGC-GAN achieves significant performance improvements in various style transfers. Although PGC-GAN excels in handling complex style transfer tasks, it tends to focus heavily on texture from style images, sometimes overlooking texture scales in content images. This issue is particularly noticeable when replacing complex image styles with those of simpler images. Therefore, in our future work, we will explore ways to enhance the performance of style transfer networks across different scales.

Declarations

All authors declare that they have no conflicts of interest.

References

Cao, P., Zhu, Z., Wang, Z., Zhu, Y. and Niu, Q. (2022) 'Applications of graph convolutional networks in computer vision', *Neural Computing and Applications*, Vol. 34, No. 16, pp.13387–13405.

Chen, H., Wang, Z., Zhang, H., Zuo, Z., Li, A., Xing, W. and Lu, D. (2021a) 'Artistic style transfer with internal-external learning and contrastive learning', *Advances in Neural Information Processing Systems*, Vol. 34, pp.26561–26573 [online] https://proceedings.neurips.cc /paper_files/paper/2021/hash/.

Chen, H., Zhao, L., Wang, Z., Zhang, H., Zuo, Z., Li, A., Xing, W. and Lu, D. (2021b) 'Dualast: Dual style-learning networks for artistic style transfer', in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.872–881.

Chiu, T.Y. and Gurari, D. (2022) 'Photowct2: compact autoencoder for photorealistic style transfer resulting from blockwise training and skip connections of high-frequency residuals', in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp.2868–2877.

Chiu, T.Y. and Gurari, D. (2023) 'Line search-based feature transformation for fast, stable, and tunable content-style control in photorealistic style transfer', in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp.249–258.

Deng, J., Dong, W., Socher, R., Li, L.J., Li, K. and Fei-Fei, L. (2009) 'Imagenet: a large-scale hierarchical image database', in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp.248–255.

Diederik, P.K. (2014) *A Method for Stochastic Optimization*, Adam, arXiv.

Farajzadeh, A., Imani, M. and Mohammadi, S. (2023) 'Hyperspectral image super resolution using anomaly weighted Gabor Based CNN', *International Journal of Information and Communication Technology Research*, Vol. 15, No. 2, pp.19–28.

Giraldo, J.H., Skianis, K., Bouwmans, T. and Malliaros, F.D. (2023) On the trade-off between over-smoothing and over-squashing in deep graph neural networks', in Proceedings of the *32nd ACM International Conference on Information and Knowledge Management*, pp.566–576.

Han, X., Wu, Y. and Wan, R. (2023) 'A method for style transfer from artistic images based on depth extraction generative adversarial network', *Applied Sciences*, Vol. 13, No. 2, p.867.

He, K., Zhang, X., Ren, S. and Sun, J. (2016) 'Deep residual learning for image recognition', in Proceedings of the *IEEE Conference on Computer Vision and Pattern Recognition*, pp.770–778.

Huang, X. and Belongie, S. (2017) Arbitrary style transfer in real-time with adaptive instance normalization', in Proceedings of the *IEEE International CONFERENCE on computer Vision*, pp.1501–1510.

Jiao, L., Chen, J., Liu, F., Yang, S., You, C., Liu, X., Li, L. and Hou, B. (2022) 'Graph representation learning meets computer vision: a survey', *IEEE Transactions on Artificial Intelligence*, Vol. 4, No. 1, pp.2–22.

Jing, Y., Mao, Y., Yang, Y., Zhan, Y., Song, M., Wang, X. and Tao, D. (2022) 'Learning graph neural networks for image style transfer', in *European Conference on Computer Vision*, pp.111–128.

Kwon, G., and Ye, J.C., 2022. Clipstyler: Image style transfer with a single text condition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 18062-18071.

Li, H. and Zhu, W. (2024) *Art Image Style Conversion Based on Multi-Scale Feature Fusion Network*.

Lin, J. and Li, R. (2024) 'Creative design of motion graphics based on BicycleGAN algorithm', *Intelligent Systems with Applications*, Vol. 21, p.200323, https://doi.org/10.1016/j.iswa.2023.200323.

Liu, T. (2023) 'A colour transfer method of interior design based on machine learning', *International Journal of Information and Communication Technology*, Vol. 22, No. 4, pp.438–455.

Ma, X. (2024) 'A comparison of art style transfer in Cycle-GAN based on different generators', in *Journal of Physics: Conference Series*, Vol. 2711, No. 1, p.12006.

Mohammad, S. and Kiritchenko, S. (2018) 'Wikiart emotions: an annotated dataset of emotions evoked by art', in Proceedings of the *Eleventh International Conference on Language Resources and Evaluation* (*LREC 2018*).

Ni, L. (2024) 'Impact analysis of informatization means driven by artificial intelligence technology on visual communication', *International Journal of Advanced Computer Science and Applications*, Vol. 15, No. 10, DOI: 10.14569/ijacsa.2024.0151055.

Ronneberger, O., Fischer, P. and Brox, T. (2015) 'U-net: convolutional networks for biomedical image segmentation', in *Medical Image Computing And Computer-Assisted Intervention-MICCAI 2015: 18th International Conference*, Munich, Germany, October 5–9, Proceedings, part III 18, pp.234–241.

Shi, T., Chen, C., Li, X. and Hao, A. (2024) 'Semantic and style based multiple reference learning for artistic and general image aesthetic assessment', *Neurocomputing*, Vol. 582, p.127434, https://doi.org/10.1016/j.neucom.2024.127434.

Tian, F. (2024) 'Denoising Monte Carlo rendering models combined with JDKPNet in artistic design application scenarios', *Intelligent Systems with Applications*, Vol. 22, p.200338, https://doi.org/10.1016/j.iswa.2024.200338.

Wang, R. and Yaermaimaiti, Y. (2024) 'Chinese character style transfer based on improved StarGAN v2 network', *International Journal of Information and Communication Technology*, Vol. 24, No. 6, pp.71–91.

Wang, S. and Li, J. (2023) 'Facial expression recognition of aerobics athletes based on CNN and HOG dual channel feature fusion', *International Journal of Information and Communication Technology*, Vol. 22, No. 3, pp.281–293.

Wang, Z., Zhang, Z., Zhao, L., Zuo, Z., Li, A., Xing, W. and Lu, D. (2022) 'AesUST: towards aesthetic-enhanced universal style transfer', in Proceedings of the *30th ACM International Conference on Multimedia*, pp.1095–1106.

Wei, Y. (2024) 'Artistic image style transfer based on CycleGAN network model', *International Journal of Image and Graphics*, Vol. 24, No. 4, p.2450049.

Xu, W., Long, C. and Nie, Y. (2023) 'Learning dynamic style kernels for artistic style transfer', in Proceedings of the *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.10083–10092.

Xu, W., Long, C., Wang, R. and Wang, G. (2021) 'Drb-GAN: a dynamic resblock generative adversarial network for artistic style transfer', in Proceedings of the *IEEE/CVF International Conference on Computer Vision*, pp.6383–6392.

Zhang, H., Lu, G., Zhan, M. and Zhang, B. (2022) 'Semi-supervised classification of graph convolutional networks with Laplacian rank constraints', *Neural Processing Letters*, pp.1–12.

Zhou, B., Lapedriza, A., Khosla, A., Oliva, A. and Torralba, A. (2017) 'Places: a 10 million image database for scene recognition', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 40, No. 6, pp.1452–`1464.