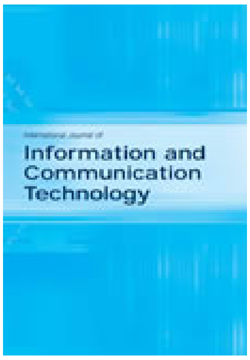


International Journal of Information and Communication Technology

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

Optimisation of art image style migration algorithms based on deep neural networks

Xiangkun Gao

DOI: 10.1504/IJICT.2025.10073121

Article History:

Received:	22 June 2025
Last revised:	14 July 2025
Accepted:	15 July 2025
Published online:	17 September 2025

Optimisation of art image style migration algorithms based on deep neural networks

Xiangkun Gao

School of Design and Art,
Jingdezhen Ceramic University,
Jingdezhen 333000, China
Email: 18179821660@163.com

Abstract: As deep learning technology has grown quickly, art image style migration, which is an important area of research in computer vision, has become more and more common. The conventional style migration technique exhibits issues such as inconsistent outcomes and diminished computing efficiency. This research presents a deep neural network (DNN)-based optimisation framework for art image style migration techniques, with the objective of optimising the style migration effect while minimising computational overhead and improving operational efficiency. The suggested algorithm demonstrates excellent migration effects and high efficiency on two real experimental datasets by optimising the network topology, integrating varied loss functions, upgrading the style expression mechanism, and improving operational efficiency. The experimental results demonstrate that the suggested algorithm surpasses existing mainstream approaches in various respects and possesses significant practical application value. Finally, the study talks about the algorithm's flaws and suggests areas for future research.

Keywords: art images; deep neural network; DNN; style migration optimisation; algorithm optimisation.

Reference to this paper should be made as follows: Gao, X. (2025) 'Optimisation of art image style migration algorithms based on deep neural networks', *Int. J. Information and Communication Technology*, Vol. 26, No. 33, pp.18–37.

Biographical notes: Xiangkun Gao received his PhD from the Jingdezhen Ceramic University in 2021. Currently, he serves as a Lecturer at the Jingdezhen Ceramic University. His research interests include design theory, ceramic art design, and artificial intelligence design.

1 Introduction

Artificial intelligence (AI) technology is changing the way people make things and run their lives at an unparalleled speed. As AI-generated content (AIGC) technology keeps becoming better, it can now make images, music, video, and other types of content in more than one way (Foo et al., 2025). This not only changes how automatic design and digital content creation work, but it also opens new possibilities for artistic creativity. In particular, the use of deep neural network (DNN) in image production greatly enhances the quality and speed of image synthesis. This allows machines to imitate artistic styles, aesthetic qualities, and human creative intent, leading to the new discipline of computational art. Art image style migration is becoming a significant tool for AI in cultural art and design innovation. It is an important part of AIGC visual generation (Wang et al., 2025).

Early style migration approaches mostly depend on conventional image processing techniques, such as image filtering, texture generation, and colour mapping. These techniques can extract and modify low-level visual information, such as edges and colour distributions, to some degree. However, it is clear that they are not good enough to capture and express high-level artistic stylistic features. It is hard to get the results you want with standard approaches, especially when it comes to showing complicated image structures and free strokes. Some algorithms are trying to use image sparse representation, wavelet transform, graph matching, and other techniques to better explain the connection between content and style as machine learning methods get stronger. Even though local texture migration has come a long way, the final outputs frequently do not look good or have a consistent style, which makes it hard to meet the needs of high-quality creative creation.

DNN has changed the way art images move through styles in a big way. The neural style migration approach may successfully separate the content and style information of an image and combine them in a natural way during the generation process. This is possible because convolutional neural networks (CNNs) are very good at extracting image features (Liu et al., 2021). This approach enables the network to capture low-level elements like colour and texture while simultaneously simulating the high-level semantic structure and spatial composition of the image, resulting in a more natural and refined migration effect.

Even though current research has produced impressive achievements, art picture style migration still has a lot of problems to solve, such as model expressiveness, style fidelity, image clarity, and conversion efficiency. The model needs to be efficient and straightforward to set up, especially for real-world use. It should also offer multi-style and multi-scale generalised migration with strong control and interactivity. There is presently an absence of DNN architecture capable of accommodating high-quality art style fusion while being adaptive and easily expandable (Al-Khazraji et al., 2023). Consequently, thorough optimisation of the existing deep style migration models in critical areas such as network architecture, loss function formulation, and feature fusion mechanisms has emerged as a significant advancement to foster further progress in this domain.

It is theoretically significant and practically valuable to explore the optimisation of DNN-based style migration methods for art images. This research can enhance the artistic expressiveness and computational efficiency of the image generation system. Additionally, it can offer innovative solutions to the fundamental challenges of multi-style modelling, style expression control, and network lightweighting, thereby broadening the application of style migration technology in content creation, cultural communication, and human-computer interaction. We want to make an image style migration method that combines artistry, accuracy, and scalability by improving the network structure, combining different loss functions, enhancing the style expression mechanism, and making the operation more efficient. This will give us better technical support for generating high-quality images.

2 Relevant technologies

2.1 Deep neural network

DNN is a neural information processing system made up of several nonlinear mapping layers. Its main principle is to use a multi-level feature transformation process to gradually turn the input data into higher-level, more semantically rich feature representations. DNN is better than traditional shallow models at fitting complicated functions, automatically finding useful patterns in large amounts of data, and doing end-to-end modelling without having to manually design features. Because of this, it is widely used in image recognition, speech analysis, natural language processing, and other areas.

For image processing tasks, DNN is frequently built on top of CNN (Borkar and Karam, 2019). The network can efficiently capture spatial local structural information in an image by using local sensing fields, weight sharing, and layer stacking. It can then create a step-by-step representation from edge texture to semantic forms in a multilayer network. Each layer in the neural network is made up of neurons that take in weighted inputs from the layer before it and use the activation function to make nonlinear outputs. This makes a nonlinear function approximator.

The output of the neural network's forward propagation can be written mathematically as follows:

$$x^{(l+1)} = \sigma\left(W^{(l)} \cdot x^{(l)} + b^{(l)}\right) \tag{1}$$

where $x^{(l)}$ is the input to layer l, $W^{(l)}$ is the weight matrix, $b^{(l)}$ is the bias term, and $\sigma(\cdot)$ is the activation function.

Every layer of the network does this action again and over again, and the last layer turns the network's high-dimensional feature mapping into real predictions. Then, during the training process, the network uses a loss function to find the difference between the output and the expected value. A backpropagation method and a gradient descent optimisation approach are then used to change the parameters (Haji and Abdulazeez, 2021). The chain rule in the neural network is unfolded layer by layer in backpropagation which makes it possible to train deep models.

DNNs are often used for difficult visual generating tasks like art picture style migration because they can model image features in a hierarchical way. The low-level network can find fundamental visual features like edges and textures. The mid-level network can put together information about local structure and shape. The high-level network can represent the overall semantic composition and spatial layout. This function is perfect for the style expression demands of art images. Style expression is not just about changing the colour or texture; it's also about how the brushstrokes, colours, and structure of the image come together to create a whole. The most common way to move art styles is to use DNN to extract multi-layer characteristics and then combine them with the style expression module (Zaman et al., 2023).

In conclusion, DNN serves as the fundamental modelling framework for the image style migration task, exhibiting robust image representation capabilities and establishing a reliable foundation for future style migration models concerning structure preservation, style reconstruction, and generation control. On this work, the enhancement of succeeding style migration algorithms will be grounded on DNN, integrated with multi-module fusion and loss mechanism guidance, to elevate overall generation quality and operational efficiency.

2.2 Migration of artistic image styles

The objective of artistic image style migration is to accomplish a seamless integration of content images and style images, resulting in synthetic images that preserve the content's structure while embodying the desired artistic style. This job has evolved through various technical iterations, transitioning from standard image processing algorithms to deep learning models, thereby establishing diverse technical schools and methodological frameworks. In general, style migration methods can be divided into groups like traditional image processing approaches, neural network-based optimisation methods, feed-forward generative models, normalisation mechanisms, and generative adversarial networks (GANs).

Most of the early approaches for migrating art styles are based on old image processing methods. They typically use colour statistics, texture synthesis, and image deconstruction to move the style first. One of the simpler types of approaches is colour histogram matching. The overall colour distribution of the content image matches that of the style image by calculating the histogram distribution of the content image and the style image in different colour channels and changing the colour of the content image (Yoo et al., 2022). This method is easy to use and works well for tone migration, but it cannot handle complicated textures and structures, which makes the migration outcomes boring and unartistic. Image pyramid and multi-scale fusion technologies work by breaking an image down into multiple resolutions and then merging them to create the final product. This method can keep the content structure intact to some degree while moving the texture's local features. However, it is hard to coordinate between different scales, which can make the image look confusing.

On the other hand, the PatchMatch algorithm and its variations use local block matching to move texture styles. The approach looks for image blocks in the content image that have a texture comparable to the stylised image. It then replaces or merges them to create the image with stylised texture. PatchMatch is better at keeping local textures intact, but it can cause structural misalignments and clear splicing traces because it does not have global semantic guidance (Bharathiraja et al., 2024). The Markov random field (MRF) model represents style migration as a probabilistic graphical model that ensures statistical consistency of local texture by establishing neighbourhood associations among pixels. The MRF method focuses on local smoothing and texture continuity (Han et al., 2020). It is quite good at preserving details, but it is very hard to process high-resolution photos in real-time since it is so complicated. Traditional methods were important in the early stages of style migration, but they have some problems. For example, they do not model high-level semantic information well enough, they cannot express styles well enough, and the quality of the results they produce is not always stable. This makes it hard to meet the needs for detail richness and structural integrity in modern art style migration.

Deep learning has played a big role in the growth of strategies for migrating artistic styles. DNN allows for the separation and reformation of high-dimensional representations of image content and style via multilayer nonlinear mapping, ushering style migration into the realm of end-to-end learning. The neural style migration method was the first to use optimisation-based style migration. The method uses pre-trained

CNNs to get multi-layer features from content and style images. It then uses the weighted combination of content loss and style loss as the objective function to improve the output images through gradient descent iterations (Gong et al., 2020). The content loss typically quantifies the Euclidean distance between the generated image and the content map within the high-level feature space, whereas the style loss assesses the feature correlation of the style map using the Gram matrix. This method can provide very high-quality style migration with rich style details and structural faithfulness, but the iterative computing cost is significant, which is challenging to match the real-time application requirements.

Normalisation methods are very significant in style migration networks. Instance normalisation (IN) is very important for style representation since it removes style information from photos by changing the statistics for each image separately, which makes the style migration effect more stable and consistent (Abdal et al., 2021). Based on this, conditional normalisation (CN) and adaptive instance normalisation (AdaIN) are improved. AdaIN makes the style conversion process more flexible by directly changing the meaning and variance of the content image features (Mulay et al., 2023). This allows for multi-style fusion and smooth control, and it encourages the growth of different style migration techniques.

By adding an adversarial training mechanism, GAN makes generated images look more genuine and stylistically expressive. In style migration, the generative network creates new images, while the discriminative network decides if the image fits the target style distribution. GAN successfully addresses the deficiencies of loss function-based style migration regarding detail representation, enhancing the naturalness and diversity of the migrated images. The GAN-based style migration method further improves unsupervised style migration, multi-style migration, and style mixing. This opens new options for making beautiful images (Shu et al., 2021). The researchers added the attention mechanism to style migration to make it easier to match content and style regions. The attention mechanism may change the network's attention weights on different spatial regions in real-time. This makes it easier to move and combine important textures and structures. The network can combine local details with the global layout to create a more realistic and delicate style impact when used with the multi-scale feature extraction technique. The transformer model has recently been integrated into the domain of style migration through its global self-attention mechanism (Du et al., 2024). This structure can capture the dependencies of the image's global scope, get beyond the problem of the typical convolutional network's limited receptive field, and make complicated artistic styles look consistent across the board. The transformer style migration model is better at migrating styles across many domains and styles over multiple domains, but its high computational cost and difficulties of training still make it hard to use in real life.

In general, art image style migration approaches are getting better at being high quality, real-time, varied, and controllable. Traditional approaches give style migration on a theoretical basis and a first step toward implementation. DNN, on the other hand, makes a big jump forward in the ability to convey style and the effect of generation.

3 Algorithm framework design and optimisation strategy

The goal of this algorithmic framework is to make art image style migration better and faster. The system has four main parts: one for building the network structure, one for integrating the loss function, one for optimising style expression, and one for making operations more efficient.

3.1 Network architecture module

This module is the most important part of the DNN-based style migration method for art images. The quality of its design has a direct impact on the model's ability to express itself, the style of migration effect, and the speed at which it works. DNN uses multi-layer nonlinear transformations to efficiently abstract and rebuild image content and style attributes. To improve the quality of style migration, it is important to build a reasonable and efficient network structure. In this paper, we utilise a deep convolutional neural network architecture founded on the Residual Network (ResNet), employing residual block as the fundamental unit (Agrawal and Govil, 2023). This approach effectively addresses the training challenges associated with gradient vanishing or gradient explosion in conventional deep networks. The main idea behind residual block is to skip some of the network layers by constantly mapping them. This way, the input feature x can be directly added to the converted residual function $F(x,\{W_i\})$ to make the output feature y:

$$y = F\left(x, \{W_i\}\right) + x \tag{2}$$

where $F(x,\{W_i\})$ is made up of things like convolutional layers, batch normalisation, activation functions, and so forth. Adding residual connectivity not only speeds up the model's convergence, but it also makes the network better at revealing complicated content and style elements, especially in deep structures, where it performs and stays stable better.

A multi-scale convolution module is added to the network topology to help the model better capture information at different levels. The module has several branches that run in parallel and use different-sized convolution kernels (3 × 3 and 5 × 5) to get local and global structural information from the image. We can combine multiscale characteristics by splicing:

$$F_{multi} = \text{Concat}\left(F_{3\times3}, F_{5\times5}\right) \tag{3}$$

This design allows the DNN to strike a good balance between keeping the details of the content and showing the style. This makes the generated image better show the rich texture of the style and the spatial relationship of the content, which greatly improves the style migration's visual effect and artistic expression.

This work presents the inverted residual block and linear bottleneck structure in the MobileNetV2 architecture to meet the real-time and resource consumption needs of style migration. Depthwise separable convolution, which breaks down ordinary convolution into two steps, makes this lightweight architecture much simpler to understand and use (Nguyen, 2020).

Also, to make the model pay even more attention to style features, the network structure includes the squeeze-and-excitation (SE) module from the channel attention mechanism (Bi et al., 2024). The SE module takes the global information from each channel and compresses it to get the channel description variable z. It then uses the activation function to do a nonlinear transformation over the two fully connected layers to get the channel weights coefficients s:

$$s = \sigma\left(W_2 \cdot \delta\left(W_1 \cdot z\right)\right) \tag{4}$$

where σ is the Sigmoid function, δ is the ReLU activation function, W_1 and W_2 are the weight matrices that can be trained. The network can change the response intensity of each channel by applying this weight coefficient to the channel attributes. This brings out vital style information and hides unnecessary features, making style migration more accurate and expressive. This attention technique not only makes it easier to send style details, but it also makes the model better at adapting to different styles.

In short, this network architecture module is based on DNN technology and includes a residual connection, multi-scale convolution, lightweight inverted residual block, and channel attention mechanism. It builds a deep convolutional neural network architecture that can express features well and run quickly, which is a good technical foundation for high-quality, real-time art image style migration.

3.2 Loss function module

This module is the most important portion of the DNN-based art picture style migration process. It is in charge of setting multi-dimensional limits and giving the created image direction. A well-thought-out design and combination of several loss processes can greatly enhance the resulting image’s content preservation, style restoration, and overall naturalness. This article examines the characteristics of artistic picture style migration and employs three principal loss functions: content loss, style loss, and total variation loss, to ensure the quality of created images from various perspectives.

Content loss is mostly utilised to guarantee that the output image faithfully preserves the spatial structure and semantic information of the input content image. The content loss is based on the feature representation of an intermediate layer in the pre-trained DNN (Chen et al., 2022). It finds the difference between the generated image and the content image in that layer’s feature space. This is how it looks in math:

$$L_{content} = \frac{1}{2}\sum_{i,j}\left(F_{ij}^{G} - F_{ij}^{C}\right)^2 \tag{5}$$

where F_{ij}^{G} and F_{ij}^{C} are the feature maps of the generated image and the content image in a certain convolutional layer, i and j are the spatial position indices. This loss function is one of the most important rules for migration work. It keeps the structural content of the image from changing too much during the process, which helps keep the visual semantics consistent.

The style loss, on the other hand, looks at how well the generated image matches the target style image in terms of brush strokes, colour, and texture. It employs gram matrix-based feature relevance matching algorithms to figure out style loss (Yu and Zheng, 2023). The style loss is the weighted total of the differences in the multi-layer Gram matrix:

$$L_{style} = \sum_{l} w_l \left\| G_l^{G} - G_l^{S} \right\|_F^2 \tag{6}$$

where G_l^{G} and G_l^{S} are the Gram matrices of the convolved features of the generated and stylised images at layer l respectively. The Frobenius paradigm is represented by $\|\cdot\|_F$, and the weight coefficients w_l control how much each layer contributes. The Gram matrix shows the overall texture and colour distribution of the image style by taking the inner product of the feature channels. The network can capture and move the complex features of the artistic style and improve the artistic expression of the resulting image by optimising this loss.

To avoid unnatural noise and artefacts in the generated image and to make the image smoother and more visually coherent, this study also presents total variation loss, which is defined as:

$$L_{tv} = \sum_{i,j}\left(\left(\hat{I}_{i,j+1} - \hat{I}_{i,j}\right)^2 + \left(\hat{I}_{i+1,j} - \hat{I}_{i,j}\right)^2\right) \tag{7}$$

This loss function enhances spatial smoothness and continuity of the image by imposing penalties on intensity disparities between adjacent pixels, mitigating stray artefacts and uneven textures resulting from style fusion, and elevating the overall image quality (Karim et al., 2023).

The content loss makes sure that the content of the generated image is kept correctly; the style loss makes sure that the target art style is moved correctly, and the total variation loss makes sure that the image is as comfortable and continuous as possible. This paper builds a multi-level and multi-angle training target system by optimising these three loss functions in a way that works together. This lets the DNN keep the content structure and show off the rich and delicate artistic style in the artistic image style migration task, while also making clear and natural high-quality images.

3.3 Stylistic expression module

This module is a fundamental part of the DNN-based art image style migration process. Its job is to efficiently extract and properly combine the target style elements so that the output images can naturally show style and have a variety of styles. The fundamental idea behind this module is to create a good way to represent style features and a good way to combine features so that the generated image may accurately show the goal style while keeping the structural information of the content image.

To begin with, this study uses the conditional instance normalising (CIN) method in the normalising methodology to accurately capture the multilevel texture and colour properties of art styles. CIN changes the meaning and variance of the feature maps in real-time by changing the style conditions' parameters (Kandula et al., 2023). This makes it possible to show different styles in a flexible way. This is the exact formula:

$$\mathrm{CIN}\left(x; \gamma_s, \beta_s\right) = \gamma_s \cdot \frac{x - \mu(x)}{\sigma(x)} + \beta_s \tag{8}$$

where x is the input feature map, $\mu(x)$ and $\sigma(x)$ are the mean and standard deviation of the feature map, γ_s and β_s are the scaling and panning parameters for the target style s. This normalisation method lets the network change the feature distribution in response to varied style conditions, which makes multi-style migration more powerful.

Second, this study creates a feature fusion module based on the self-attention mechanism to make sure that content and style features work well together. The self-attention mechanism can dynamically capture the long-range dependencies between different parts of the image (Yang et al., 2021). This helps the network match the content structure and style texture more correctly. The main way it works is:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^{\top}}{\sqrt{d_k}}\right)V \tag{9}$$

where Q, K, and V are the query, key, and value matrices, and d_k is the size of the key vector. The self-attention method improves the expression of essential style regions while keeping the content's spatial coherence by weighting and summing the feature maps. This makes the style of migration impact more natural and richer.

The style expression module combines the CIN and the self-attention method. This not only makes multi-style migration more flexible and accurate, but it also makes style features better at modelling spatial dependencies. The module accomplishes in-depth expression and successful fusion of artistic styles by dynamically modifying the feature distribution and recording the global feature associations. This greatly improves the artistic expression and visual effect of the images it creates.

3.4 Efficiency optimisation module

This module's goal is to fix the DNN's inference efficiency problem in art picture style migration and make the model run faster and use fewer resources in real-world situations. The style migration algorithm frequently necessitates frame-by-frame processing at the image level; thus, excessive computing redundancy complicates the fulfilment of requirements for real-time creation, mobile deployment, and other contexts. Thus, this study initiates the development of an optimisation system by focusing on two key areas: network lightweight modelling and model compression. This system aims to achieve economic operational capabilities while preserving the quality of style migration.

This study presents depthwise separable convolution as a substitute for the conventional convolution process, breaking it down into depthwise convolution and pointwise convolution, which cuts down on the amount of work that needs to be done (Shan et al., 2020). The scales of the parameters are compared below:

$$\frac{C_{in} \cdot K \cdot K + C_{in} \cdot C_{out}}{C_{in} \cdot C_{out} \cdot K \cdot K} = \frac{1}{C_{out}} + \frac{1}{K^2} \tag{10}$$

where K is the size of the convolutional kernel, C_{in} and C_{out} are the number of input and output channels, respectively. When C_{out} and K are big, you can save roughly 90% or more on computing costs. This method makes the network far less complex and memory-intensive during the inference phase without losing much of its ability to express itself.

This study also talks about the knowledge distillation approach, which helps make the model smaller and run more efficiently (Gou et al., 2021). By training a lightweight student network to meet the output features of a more powerful but bulky teacher network, this strategy cuts down on the size of the model while keeping its performance high. The loss function for knowledge distillation is:

$$L_{KD} = \alpha \cdot L_{CE}\left(y_s, y_{gt}\right) + (1-\alpha) \cdot L_{KL}\left(y_s, y_t\right) \tag{11}$$

where L_{CE} is the cross-entropy loss between the student output y_s and the actual label y_{gt}, L_{KL} is the Kullback-Leibler scatter between the student output and the teacher output y_t, α is the weight balance factor between the two. This way, the student network not only learns the task aim, but also picks up on the style expression nuances and discriminative information in the instructor network. This makes it better at generalising and migrating.

In conclusion, the efficiency optimisation module allows the DNN to significantly compress parameters and speed up processing while keeping the quality of art image style migration by using advanced techniques like depth-separable convolution and knowledge distillation. This meets the overall needs of real-time, multi-device adaptation and low-energy computing, and gives the algorithm a strong operational guarantee for its wide use.

In summary, the DNN-based art image style migration algorithm framework provided in this paper systematically designs and improves four areas: network structure, loss function, style expression, and efficiency optimisation. First, ResNet, multi-scale convolution, and an attention mechanism are used to build a DNN structure with strong expressive power, which is a basic requirement for high-quality image generation. Second, multiple objectives, such as content loss, style loss, and total variation loss, are combined to accurately guide the model to learn the artistic features. At the level of style expression, the combination of CN and a self-attention mechanism achieves the dynamic modelling and spatial fusion of stylistic features. Finally, the system is designed and improved by depth-separable loss function, style expression, and efficiency optimisation. The combination of CN and self-attention mechanism achieves dynamic modelling and spatial fusion of style features. Finally, the strategies of deep separable convolution and knowledge distillation effectively improve the model's inference efficiency and deployment flexibility. The four modules work together to make a style migration algorithm system that is both artistically expressive and useful in real life.

This study creates the following pseudo-code framework (Algorithm 1) to make it easier to see how the suggested algorithm works as a whole and how each module works together to improve the overall process.

Algorithm 1 Pseudo-code for the proposed DNN-based style transfer algorithm

Input: Content images, style images, initial DNN weights, learning rate, total iterations

Output: Optimised DNN weights

```
begin
    Initialise DNN weights and lightweight network structure
    for iteration = 1 to total_iterations do
        Load a batch of content and style images
        // Module 1: Network structure
        Extract content features using DNN encoder
        Extract style features using DNN encoder
        Fuse content and style features to generate combined features
        Generate output image from combined features via decoder

        // Module 2: Loss function
        Calculate content loss between output and content features
```

```
        Calculate style loss between output and style features
        Calculate structure loss to preserve image structure
        Combine losses into total loss

        // Module 3: Style expression
        Extract style embedding from style features
        Enhance combined features with style embedding
        Update output image accordingly

        // Module 4: Efficiency optimisation
        If inference time is too high or frame rate too low then
            Apply model pruning and quantisation
            Update network weights accordingly
        end if

        Backpropagate total loss and update DNN weights
    end for
    return optimised DNN weights
end
```

4 Design of experiment and analysis of results

4.1 Experimental environment and dataset

To validate the efficacy of the proposed DNN-based optimisation method for the art image style migration algorithm, this paper undertakes comprehensive experimental evaluations on publicly accessible art image datasets and executes model training and testing on a conventional deep learning computing platform.

The hardware environment for the experimental platform consists of an Intel Core i9-12900K processor, an NVIDIA RTX 3090 graphics card (with 24 GB of video memory), 128 GB of DDR5 RAM, and the operating system is Ubuntu 22.04. The tests are carried out in a software environment that uses PyTorch 2.0 as the main deep learning framework. The main deep learning framework for the software environment is PyTorch 2.0, the version of Python is 3.10, and the supporting libraries include numpy, matplotlib, and others. The model is trained with the optimiser during the model training procedure. The optimiser is set to Adam, the initial learning rate is set to 0.0001, the batch size is set to 8, and the training images are all scaled to 256×256 pixels.

For dataset construction, this study uses two real art picture datasets: the WikiArt dataset and the Painter by Numbers dataset. The WikiArt dataset is used as the style image source, while the painter by numbers dataset is utilised as the content image source to make the image pairs for the style migration task. Both are commonly utilised in research activities including modelling image styles, analysing artwork, and migrating styles. They are both very useful and representative. Table 1 shows the basic facts about the dataset.

Table 1 Overview of datasets used in the style transfer experiment

Dataset name	*Number of images*	*Image type*	*Primary use*
WikiArt dataset	80,000	Oil paintings, sketches, watercolours	Style reference
Painter by numbers	100,000	Digital paintings, artworks	Content input

All images are uniformly normalised (256×256) prior to model input. Lightweight data enhancement techniques, such as random cropping, mirror flipping, and colour perturbation, are employed to improve the model's robustness and generalisation during training.

This paper creates a multi-style, multi-subject, multi-author image pair combination from the dataset above. This combination is useful for training and testing the style migration algorithm in a real art image environment and gives a solid database for the performance test that will follow.

4.2 Experimental indicators

This paper establishes a multi-dimensional quantitative assessment system using three classical and representative objective evaluation indices to thoroughly assess the performance of the proposed DNN-based art image style migration algorithm concerning content retention, style migration effect, and image structural integrity.

Content similarity (CS) is the first way to quantify how well content is kept. This metric uses the pre-trained VGG network to find the mid-level characteristics of the content image and the generated image, then it finds the Euclidean distance or cosine similarity between the two. This shows how well the image created keeps the structure of the original content. The higher the value of this metric (or the lower the Euclidean distance), the better the content information is kept. The exact range of values depends on how the metric is calculated.

Second, to find out how style migration affects things, we employ style similarity (SS) based on the Gram matrix. This metric shows how far apart the two are in the texture and style feature space by finding the difference between the Gram matrices of several convolutional layers in the generated image and the target style image. Better style migration happens with smaller values, which can be any non-negative real integer and do not have to be strict.

To make sure that the photos look natural and are structurally sound, we also use the well-known structural similarity index (SSIM). SSIM uses brightness, contrast, and structure to compare the two images and give them a score between 0 and 1. A score closer to 1 means that the structure is more accurate and less distorted.

In short, the three metrics CS, SS, and SSIM test the style migration algorithm's performance in terms of semantic content, artistic style, and image structural integrity. They also show the strengths and weaknesses of the algorithms in real-world tasks in a more complete and objective way. The following experimental section will quantitatively evaluate and contrast this paper's methodology with several comparison approaches grounded in the indicators.

4.3 Comparative experiment on style migration effect

This section develops style migration effect comparison experiments to fully evaluate the migration quality and performance capabilities of the DNN-based art picture style migration algorithm described in this study. The experiment seeks to validate the proposed algorithm's superiority over conventional style migration techniques regarding content retention, style expression, and structural fidelity, hence demonstrating the algorithm's practical significance and novel impact.

This experiment utilises two public art image datasets, each offering a diverse array of art styles and content images. The WikiArt dataset encompasses numerous renowned art genres and painting styles, making it ideal for assessing the model's performance in intricate style scenarios. Conversely, the painter by numbers dataset emphasises a broad spectrum of painting subjects and texture characteristics, facilitating the evaluation of the model's generalisation capability and the impact of detail migration. The consequence of detail migration. Everyone agreed on using CS, SS, and SSIM as the evaluation indexes.

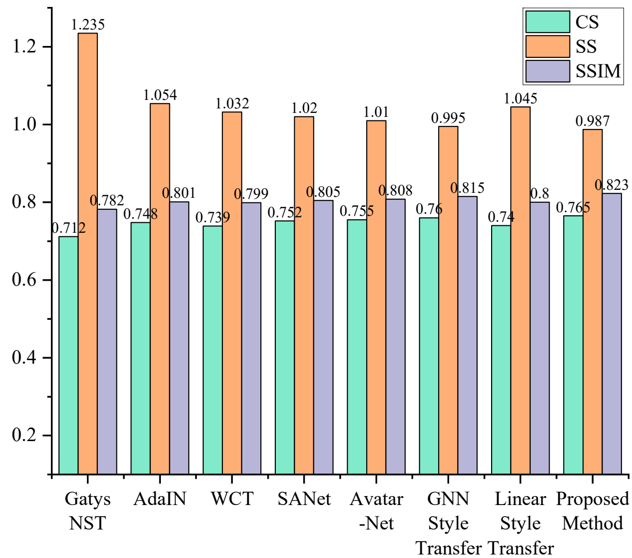
To demonstrate the thoroughness of the comparison, seven representative and commonly utilised models in the domain of style migration have been chosen, including:

- Gatys neural style transfer (Gatys NST): the first optimal style transfer approach that uses deep CNNs. It makes images by repeatedly optimising content and style loss, which works well but takes a long time.
- Adaptive instance normalisation*:* it is a new way to change styles that work quickly, with several styles, and does not require training by changing the instance normalisation parameter to find the right balance between speed and effect.
- Whitening and colouring transform (WCT): it is based on the features of space whitening and colouring transformation. It makes it easier to change style features and works better for style migration.
- Style-attentional network (SANet): it uses the attention mechanism to dynamically allocate weights during the fusion of style and content aspects, which improves the quality of detail migration.
- Avatar-net: it uses a self-attention mechanism and multi-layer feature matching to make it easier to move between different styles and better show the details of artistic style.
- GNN style transfer: it uses graph neural networks to understand how the structure of a picture and its style aspects are related in a complicated way. This improves the retention of content structure and the expression of style.
- Linear style transfer: a quick way to change styles using linear transformation that works best in contexts with limited resources.

Figure 1 compares how well different models work on the WikiArt dataset.

The figure shows that the proposed method gets a CS score of 0.765, which is far higher than all the other models that were compared. GNN Style Transfer comes in second with a score of 0.760, while AdaIN and Avatar-Net come in at 0.748 and 0.755, respectively. This shows that the method in this study is quite good at keeping the meaning and details of the content consistent. This shows that the optimised DNN architecture can better keep the fundamental information of the content photos, lower the amount of content loss, and make the generated images more similar to the original content.

Figure 1 Performance comparison on WikiArt dataset (see online version for colours)



This paper's method does just as well on SS metrics, with a minimum value of 0.987 (lower means better style migration). This is better than 0.995 for GNN Style Transfer and 1.010 for Avatar-Net. This means that this paper's model is better at capturing the target art style features, and the style expression is more subtle and natural. The optimised model exhibits a big improvement in SS enhancement compared to the traditional Gatys NST (1.235). This means that it has a better ability to rebuild styles.

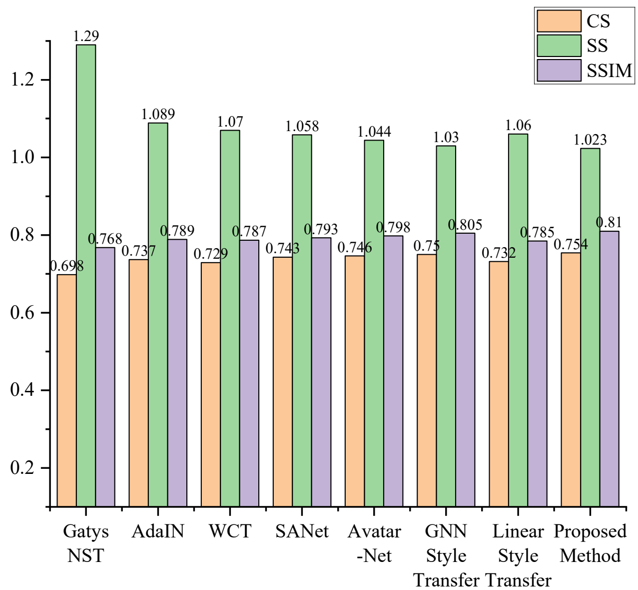
This paper's technique also comes out on top in the SSIM measure, with a score of 0.823, which is higher than GNN style transfer (0.815) and SANet (0.805). This shows that the algorithm does a good job of keeping the image's general structure and detailed texture. The created image not only fits the style, but it also has a complete structure and a more natural and coherent look. The optimised DNN architecture achieves a balanced improvement in content retention, style migration, and structural fidelity, showing great overall performance.

Figure 2 shows how the painter by numbers dataset performed compared to others.

The graphic shows that the suggested approach gets a CS metric of 0.754 on the painter by numbers dataset, which is better than all the other models that were compared. GNN style transfer has a CS value of 0.750, Avatar-Net has a CS value of 0.746, and AdaIN has a CS value of 0.737. These are the next best performers. This approach in this research greatly enhances the capacity to maintain content compared to the old Gatys NST (0.698). This means that the optimised DNN can better preserve the content image's semantic structure and detailed details, which reduces content loss during style migration.

This paper's technique also does well on SS metrics, getting a value of 1.023, which is better than GNN style transfer's 1.030 and Avatar-Net's 1.044. The SS measures have dropped significantly from Gatys NST's 1.290. This shows that the algorithm in this study is better at capturing and replicating the target art style, and the style expression is more natural and delicate. The style expression is more natural and delicate, which makes the creative effect of style migration better.

Figure 2 Performance comparison on painter by numbers dataset (see online version for colours)



The algorithm has the highest SSIM value of 0.810, followed by GNN style transfer with 0.805 and Avatar-Net with 0.798. This shows that the optimised algorithm does a good job of keeping the structural integrity and texture details of the generated images, making sure that the images still look good and feel real after the style migration cohesion and realism after style migration.

The DNN-based style migration method proposed in this paper achieves a balanced improvement in content preservation, style expression, and structural fidelity on the painter by numbers dataset. This shows that the algorithm can work with a wide range of works of art and helps create high-quality art images.

4.4 Operational efficiency testing

To further confirm the practical performance of the proposed DNN-based art image style migration algorithm in various application scenarios, this section devises efficiency test experiments to assess the algorithm's running speed, processing capacity, and model lightweight level across different image datasets, as well as to evaluate its stability and adaptability to diverse inputs.

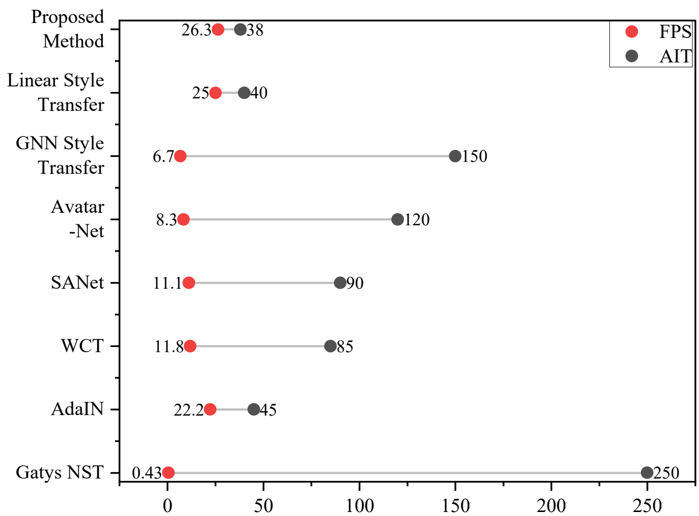
In the test, all models use the same input data and pre-processing. To make sure the measurements are accurate and can be compared, the gradient computation is switched off during the inference step. Several repeated tests are used to find the average inference time, and a system monitoring tool keeps track of how much memory is being used.

The key evaluation metrics are the following two:

- Average inference time (AIT): the unit is milliseconds, which shows how long it takes the model to process one image on average. This is a significant measure of how well the model works in real-time and how quickly it responds. The smaller the value, the better the real-time performance.
- Frames per second (FPS): AIT uses this figure to show how well the model works in real-time situations like video streaming.

Figures 3 and 4 show how well the proposed technique and six popular style migration models work on GPUs and CPUs, respectively.

Figure 3 Efficiency comparison on WikiArt dataset (see online version for colours)

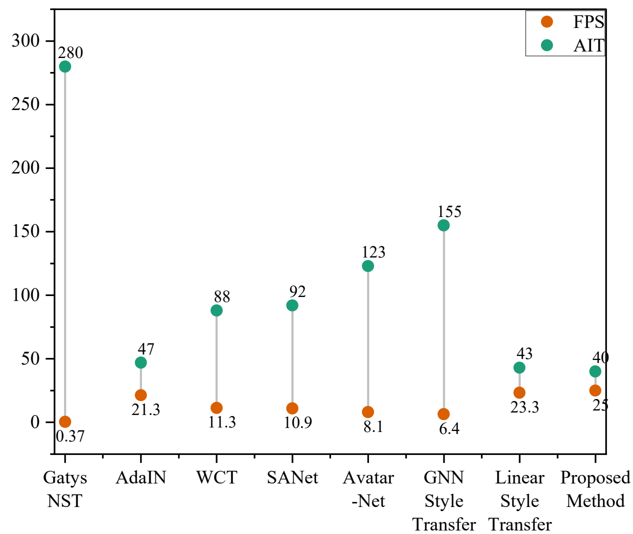


The WikiArt dataset shows that the model in this work is the most efficient overall: its AIT of 38 ms is the lowest of all models, 2 ms lower than linear style transfer (40 ms), and 212 ms faster than the classic Gatys NST (250 ms). The FPS, on the other hand, is 26.3, which is the highest of all the comparisons. The FPS, on the other hand, reaches 26.3, which is the highest of all the comparison models. This means that more than 26 photos can be processed in one second. This paper's technique has an FPS of 18.6, which is more than twice as fast as Avatar-Net's (8.3), SANet's (11.1), and WCT's (11.8). This shows how well it works for processing complicated stylised images.

AdaIN (AIT of 45 ms, FPS of 22.2) and linear style transfer (AIT of 40 ms, FPS of 25.0) are both more efficient than the method in this work, although they are still not as good. AdaIN has 7 ms longer AIT and 4.1 less FPS than this paper. Linear style transfer is similar, but it often has problems with distortion when it comes to keeping style nuances. This paper's solution increases efficiency while keeping accuracy, which shows that the optimisation strategy works.

The method still works well with the painter by numbers dataset, with an AIT of 40 ms and an FPS of 25.0. This is only a small change from the WikiArt dataset (the AIT goes up by 2 ms and the FPS goes down by 1.3). This shows that the model is still very stable in data with more complex content distribution. This paper's technique is 115 ms faster than GNN style transfer (AIT of 155 ms, FPS of 6.4), and the frame rate is almost four times higher. This further proves that it can analyse data quickly and adapt well.

Figure 4 Efficiency comparison on painter by numbers dataset (see online version for colours)



This paper's algorithm is better than SANet and WCT on painter by numbers. The AIT for SANet is 92 ms, which is over 2.3 times that of this paper, and the FPS is just 10.9. The AIT for WCT is 88 ms, and the FPS is 11.3, which is half the speed of this paper. This means that they still have a style hierarchy, but their inference speed is 115 ms faster and their frame rate is almost 6.4 faster. This demonstrates that while they possess certain advantages in maintaining the style hierarchy, fulfilling the requirements for real-time generation in actual applications proves challenging. Conversely, the method presented in this research attains a superior equilibrium between the two.

AdaIN and linear style transfer are still the most efficient, with AITs of 47 ms and 43 ms on painter by numbers and FPSs of 21.3 and 23.3, which are a little lower than those on WikiArt. The model in this study, on the other hand, has an FPS of 25.0, which is still the best. This paper's solution is more resilient when dealing with information that has complicated visual structures and big changes in texture since it works consistently across datasets.

The DNN-based art image style migration approach described in this research achieves very low AIT (38 ms and 40 ms, respectively) and high FPS (26.3 and 25.0) on both datasets. It beats all other common comparison methods. The designed network structure optimisation and module integration techniques function well to improve efficiency on datasets with diverse styles and content structures. This gives a trustworthy guarantee for real-time creation and edge deployment.

5 Conclusions

This paper proposes a DNN-based optimisation algorithm framework addressing the efficiency and effectiveness challenges in art image style migration. To fix the problems with traditional methods that make style control unstable, generation quality low, and computation costs high, this paper changes the algorithm structure in four ways: it uses a compact convolutional backbone network to speed up image feature extraction; it creates a multiple loss function that combines content and style constraints to improve migration accuracy; it adds a style expression module to improve style separation and reconstruction; and finally, it adds a feature compression and forward computation acceleration mechanism to lower computation costs. Finally, to improve efficiency, procedures for feature compression and forward computation acceleration are included to lower the cost of calculation. The experimental portion utilises two datasets, WikiArt and painter by numbers, to compare various mainstream models regarding style restoration efficacy and operational efficiency. The findings indicate that the algorithm presented in this paper surpasses existing methods across several evaluation metrics, demonstrating enhanced practicality and stability.

Nonetheless, the research presented in this publication exhibits certain limitations. The model structure finds a good mix between efficiency and migration quality, but the style fidelity still suffers when it comes to extreme styles (like abstractionism and expressionism) or photos with a lot of detail. This is because the edge regions get blurred and the brushstrokes do not seem right. Secondly, the algorithm training process needs a lot of high-quality image pairs, which makes the training cost high. Also, even though the model is lightweight, it needs to be compressed more, and the speed of the calculations needs to be improved to work with real-time interactions on edge devices with limited resources.

To tackle the challenges, subsequent research may be conducted on the following avenues:

1 Multi-style joint modelling: adding a unified style embedding space to handle numerous style migration with a single model and make the model more flexible.

2 Controllability enhancement: include parameters that users can change to allow for real-time control of migration elements like style strength and texture range and make the experience more interactive.

3 Cross-modal style migration: increasing the migration of picture styles to other modalities like speech, video, text, etc. to find more interesting ways to express style.

In summary, this study provides an optimisation approach for art image style migration that considers generation quality and computing efficiency and validates its effectiveness and adaptability by rigorous experimentation. Future efforts will focus on enhancing the algorithm's flexibility and interpretability, while also advancing the widespread implementation of style migration technology in digital art creation, cultural communication, and human-computer interaction.

Acknowledgements

This work is supported by the 2024 Annual Key Research Base Project of Philosophy and Social Sciences in Jiangxi Province (No. 24ZXSKJD38).

Declarations

All authors declare that they have no conflicts of interest.

References

Abdal, R., Zhu, P., Mitra, N.J. and Wonka, P. (2021) 'Styleflow: attribute-conditioned exploration of stylegan-generated images using conditional continuous normalizing flows', *ACM Transactions on Graphics (ToG)*, Vol. 40, No. 3, pp.1–21.

Agrawal, N. and Govil, H. (2023) 'A deep residual convolutional neural network for mineral classification', *Advances in Space Research*, Vol. 71, No. 8, pp.3186–3202.

Al-Khazraji, L.R., Abbas, A.R., Jamil, A.S. and Hussain, A.J. (2023) 'A hybrid artistic model using deepy-dream model and multiple convolutional neural networks architectures', *IEEE Access*, Vol. 11, pp.101443–101459.

Bharathiraja, S., Rajesh Kanna, B., Geetha, S. and Anusooya, G. (2024) 'Unmasking the digital deception – a comprehensive survey on image forgery and detection techniques', *Australian Journal of Forensic Sciences*, Vol. 56, No. 6, pp.635–683.

Bi, Z., Sun, S., Zhang, W. and Shan, M. (2024) 'Click-through rate prediction model based on graph networks and feature squeeze-and-excitation mechanism', *International Journal of Web Information Systems*, Vol. 20, No. 4, pp.341–357.

Borkar, T.S. and Karam, L.J. (2019) 'DeepCorrect: correcting DNN models against image distortions', *IEEE Transactions on Image Processing*, Vol. 28, No. 12, pp.6022–6034.

Chen, J., Li, X., Li, Y., Li, P., Wang, M., Zhang, X., Gong, Z., Wu, K. and Leung, V.C. (2022) 'A simple yet effective layered loss for pre-training of network embedding', *IEEE Transactions on Network Science and Engineering*, Vol. 9, No. 3, pp.1827–1837.

Du, X., Jia, N. and Du, H. (2024) 'FST-OAM: a fast style transfer model using optimized self-attention mechanism', *Signal, Image and Video Processing*, Vol. 18, No. 5, pp.4191–4203.

Foo, L.G., Rahmani, H. and Liu, J. (2025) 'AI-generated content (AIGC) for various data modalities: A survey', *ACM Computing Surveys*, Vol. 57, No. 9, pp.1–66.

Gong, D., Zhang, Z., Shi, Q., Van Den Hengel, A., Shen, C. and Zhang, Y. (2020) 'Learning deep gradient descent optimization for image deconvolution', *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 31, No. 12, pp.5468–5482.

Gou, J., Yu, B., Maybank, S.J. and Tao, D. (2021) 'Knowledge distillation: a survey', *International Journal of Computer Vision*, Vol. 129, No. 6, pp.1789–1819.

Haji, S.H. and Abdulazeez, A.M. (2021) 'Comparison of optimization techniques based on gradient descent algorithm: a review', *PalArch's Journal of Archaeology of Egypt/Egyptology*, Vol. 18, No. 4, pp.2715–2743.

Han, L., Gu, S., Zhong, D., Quan, S. and Fang, L. (2020) 'Real-time globally consistent dense 3D reconstruction with online texturing', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 44, No. 3, pp.1519–1533.

Kandula, P., Suin, M. and Rajagopalan, A. (2023) 'Illumination-adaptive unpaired low-light enhancement', *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 33, No. 8, pp.3726–3736.

Karim, S., Tong, G., Li, J., Qadir, A., Farooq, U. and Yu, Y. (2023) 'Current advances and future perspectives of image fusion: a comprehensive review', *Information Fusion*, Vol. 90, pp.185–217.

Liu, Y., Pu, H. and Sun, D-W. (2021) 'Efficient extraction of deep image features using convolutional neural network (CNN) for applications in detecting and analysing complex food matrices', *Trends in Food Science & Technology*, Vol. 113, pp.193–204.

Mulay, S., Ram, K. and Sivaprakasam, M. (2023) 'Attention adaptive instance normalization style transfer for vascular segmentation using deep learning', *Applied Intelligence*, Vol. 53, No. 24, pp.29638–29655.

Nguyen, H. (2020) 'A lightweight and efficient deep convolutional neural network based on depthwise dilated separable convolution', *Journal of Theoretical and Applied Information Technology*, Vol. 98, No. 15, pp.2937–2947.

Shan, W., Yang, M., Wang, T., Lu, Y., Cai, H., Zhu, L., Xu, J., Wu, C., Shi, L. and Yang, J. (2020) 'A 510-nW wake-up keyword-spotting chip using serial-FFT-based MFCC and binarized depthwise separable CNN in 28-nm CMOS', *IEEE Journal of Solid-State Circuits*, Vol. 56, No. 1, pp.151–164.

Shu, Y., Yi, R., Xia, M., Ye, Z., Zhao, W., Chen, Y., Lai, Y-K. and Liu, Y-J. (2021) 'Gan-based multi-style photo cartoonization', *IEEE Transactions on Visualization and Computer Graphics*, Vol. 28, No. 10, pp.3376–3390.

Wang, B., Chen, Q. and Wang, Z. (2025) 'Diffusion-based visual art creation: a survey and new perspectives', *ACM Computing Surveys*, Vol. 57, No. 10, pp.1–37.

Yang, J., Li, C., Zhang, P., Dai, X., Xiao, B., Yuan, L. and Gao, J. (2021) 'Focal attention for long-range interactions in vision transformers', *Advances in Neural Information Processing Systems*, Vol. 34, pp.30008–30022.

Yoo, W.S., Kang, K., Kim, J.G. and Yoo, Y. (2022) 'Extraction of color information and visualization of color differences between digital images through pixel-by-pixel color-difference mapping', *Heritage*, Vol. 5, No. 4, pp.3923–3945.

Yu, L. and Zheng, Q. (2023) 'AI-enhanced digital creativity design: content-style alignment for image stylization', *IEEE Access*, Vol. 11, pp.143964–143979.

Zaman, K., Sah, M., Direkoglu, C. and Unoki, M. (2023) 'A survey of audio classification using deep learning', *IEEE Access*, Vol. 11, pp.106620–106649.