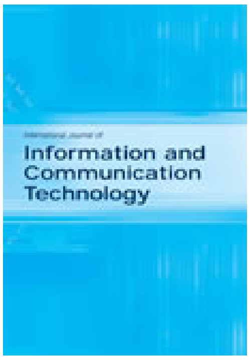


International Journal of Information and Communication Technology

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

Research on artistic style recognition and image transfer method based on deep visual feature extraction

Wei Wang, Hanchen Li, Rushana Sulaiman @ Abd Rahim, Peilei Cui

DOI: 10.1504/IJICT.2025.10074385

Article History:

Received:	02 August 2025
Last revised:	30 August 2025
Accepted:	01 September 2025
Published online:	13 November 2025

Research on artistic style recognition and image transfer method based on deep visual feature extraction

Wei Wang

College of Creative Arts,
University of Malaya,
Kuala Lumpur, 50603, Malaya, Malaysia
Email: 13593117596@163.com

Hanchen Li* and
Rushana Sulaiman @ Abd Rahim

College of Creative Arts,
University of Technology MARA,
Shah Alam, Selangor, 40450, Malaya, Malaysia
Email: li_hanchen@163.com
Email: rushana@uitm.edu.my
*Corresponding author

Peilei Cui

SophNet,
Beijing, 10000, China
Email: 15801276191@163.com

Abstract: This research presents a comprehensive approach to artistic style recognition and image style transfer using deep visual feature extraction techniques. To enhance the identification of fine art forms, the study employs a two-stage classification model that combines shallow and deep neural networks, utilising convolutional neural networks (CNNs), namely VGG16 and VGG19. A novel neural style transfer network is proposed, incorporating a coarse-to-fine methodology and whitening and colouring transformation (WCT) to preserve global content structures while effectively applying local stylistic elements. Extensive experiments on the Wiki Art and Pandora 18K datasets validate the model's ability to enhance style classification accuracy, minimise restructuring loss, and reduce runtime. The outcomes show that the suggested approach greatly improves automated art analysis and digital creative apps while keeping high-resolution image integrity and successfully integrating the unique visual traits of artistic styles.

Keywords: artistic style recognition; image style transfer; deep learning; convolutional neural networks; neural style transfer; whitening and colouring transformation; WCT; shallow neural network; SNN; feature extraction; image processing; coarse-to-fine stylisation; computer vision; machine learning; Wiki Art.

Reference to this paper should be made as follows: Wang, W., Li, H., @ Abd Rahim, R.S. and Cui, P. (2025) 'Research on artistic style recognition and image transfer method based on deep visual feature extraction', *Int. J. Information and Communication Technology*, Vol. 26, No. 40, pp.86–103.

Biographical notes: Wei Wang is currently a researcher at the College of Creative Arts, University of Malaya, Kuala Lumpur, Malaysia. His academic interests include computer vision, artificial intelligence, and digital image processing, with a focus on artistic style recognition and image synthesis. He has contributed to several interdisciplinary studies exploring the integration of technology and creative design.

Hanchen Li is a faculty member at the College of Creative Arts, Universiti Teknologi MARA (UiTM), Shah Alam, Malaysia. His research expertise lies in digital media art, deep learning applications in visual creativity, and AI-assisted design. He has published multiple works in international journals on computer graphics and intelligent art generation

Rushana Sulaiman @ Abd Rahim is affiliated with the College of Creative Arts, Universiti Teknologi MARA (UiTM), Shah Alam, Malaysia. Her research focuses on creative computing, multimedia technology, and the intersection of art and artificial intelligence. She has extensive experience in higher education and contributes to advancing innovation in digital visual arts.

Peilei Cui is currently employed at SophNet, Beijing, China. Her professional background includes research and development in intelligent systems and computer vision. Her work involves integrating neural networks into visual data interpretation, with special emphasis on enhancing digital art analysis and style transfer technologies.

1 Introduction

The most pressing problem in machine image recovery and identification is artistic style recognition, which is the focus of this study (Wang et al., 2021). To put it simply, picture artistic style detection is like assigning a category to an image. One way to categorise photos is by the objects they depict; this is achieved through the use of labels. Recognising scenes as opposed to representations, dogs as opposed to cats, and various types of written characters are all examples of object recognition in action. The second kind of image recognition is creative style recognition, which relies on the meaning of the picture to determine whether an image is aesthetically beautiful or ugly, or whether a road scenario is safe or dangerous. Recognising creative styles is difficult since it is subjective and varies from person to person, yet traditional machine learning methods work well for object recognition (Yunfei et al., 2020). It has long been a problem for computer scientists and art historians to categorise various forms of great art (Li et al., 2020). Experts often mention the artistic style recognition as an issue with this work. This expression perfectly encapsulates the gulf that exists between the limitations of traditional methods of art analysis and categorisation and the richness and complexity of creative forms. Since artistic expression is nebulous and often subjective, the gap between the two camps in terms of style recognition serves to emphasise how challenging it is to provide a specific definition and classification. Using machine learning for art classification, this research aims to solve the issue of identifying artistic styles (Ma et al., 2020). Not only does this effort finally put an end to the never-ending struggle to categorise different kinds of art, but it also ushers in a slew of groundbreaking new possibilities.

Expert art appraisers can discern a certain style in paintings because they have honed their skills over many years studying the finer points of fine art. Because they need much practice and exposure to visual stimuli, these abilities retain their secluded nature for a long time. Fine art is now more accessible than ever before thanks to the proliferation of internet collections and other digital resources. Because of this, there is a need to disseminate art knowledge to more people. Transferring human expertise to machines is one way to meet this demand. It is possible to teach computers to recognise the creative style of previously unidentified photos by training them on large databases of great art that have been annotated by humans (Benzon et al., 2022). Using this, we can automate image retrieval, categorise images by category, and classify unlabelled museum photos. Possible additional uses for machine-based art knowledge include re-discovering lost artworks and creating robots with human-like aesthetic sensitivities and an enthusiasm for art. The majority of art classification systems use style as their primary criterion for identifying and categorising paintings. In the visual arts, a style is a collection of characteristics that are associated with a certain school of thought, period, or aesthetic tradition.

The unique stylistic category of a painting might be difficult to classify, even for specialists (Bica et al., 2023). Some of the challenges include the fact that distinct and fashionable elements are not always easy to understand, subtleties that distinguish between many types of art, seamless transitions between different eras of art, and creative traits that exhibit characteristics of multiple styles or are inherently undefined. The solution to these problems, according to research, is to train the model to ignore irrelevant features and instead concentrate on those that are causally related to the labels in the samples (Li et al., 2022). While it's true that many current deep learning algorithms can learn everything from start to finish without using features, feature selection and sample representation are still necessary for tasks like pattern recognition and behaviour feature extraction. Infer that the conditional distribution of the target variable will remain unchanged when considering all direct factors impacting it, even while controlling for all other variables (Ranftl et al., 2020). For this reason, they came up with the invariant causal prediction. This was the first attempt of its kind to use 'invariance' to figure out causal structures and give a subset of the real causal factors (Imran et al., 2023).

All of these methods have been tested and shown to work well for both out-of-distribution (OOD) transfer and learning quickly. But the environment names that have already been set for the collection are what make making these models work. It is common for a model to learn invariance faster when it is put in better settings (Zhao et al., 2022). A framework for invariant learning environment inference was suggested. This framework uses pre-existing biased models to infer environments and then uses the data to assign labels to those environments. One drawback of invariant representations for environment partitioning is that they cause features associated with the environment to be progressively discarded. For complicated datasets in particular, this discarding renders possible environmental cues invisible, making it harder for the model to differentiate between input from various latent environments (Ma et al., 2020). We provide a human resource management system that can adapt to different environments by learning new invariant features, which it can then use to create fake features, which it can then use to label the environment. Drawings, the first form of creative expression, maintain the unique thoughts and emotions of its creators and are an important part of human history (Ding et al., 2024). Every great work of art carries the artist's own style and expression.

Learning what makes this style unique can help artists develop their own creative skills. So, alongside more conventional forms of art theory education, computer vision and image processing are attracting increasing amounts of focus because to the widespread adoption and rapid advancement of computing power.

2 Related work

2.1 Image style transfer

People who work in computer vision learn how to copy picture styles a lot. To do this, you need to use the style of one picture while keeping the structure of the picture's information (Li et al., 2021). At the start of the deep learning era, however, popular ways of making moving pictures started to use feed-forward networks and repeated methods to build their models. When CNNs were first used for style transfer, they had to be used over and over to make styling better by getting rid of noise in the pictures. When you use flow-based art flow and projected flow network (PFN) (Han et al., 2023), you get results that are fair. IEST and CAST, on the other hand, use different ways of learning to give you good results. Western oil painting styles have come a long way and are now used in a wider range of style. However, it is still hard to find the best way to combine them to make better art. This is because the way oil paintings are made in the West and China is very different. Someone said that the style should become more like old Chinese pen art because of this. It's hard to use modern techniques on Chinese Wuhu Iron Paintings because they have more complicated techniques and layers that are three-dimensional.

2.2 Literature review

IST has been a hotspot for AI and computer vision research since it involves rethinking an image's content within the stylistic constraints of another. Both domestically and internationally, IST techniques abound. Generally speaking, there are two main schools of thought when it comes to IST, with different theoretical foundations and application methods: classical IST and neural network-based IST (Wang et al., 2023). For instance, IST founded on the concept of texture synthesis, image filtering, stroke-based rendering (SBR), and picture analogy are all examples of traditional approaches that prioritise mathematical optimisation and human feature extraction (Zhou et al., 2024). Although these conventional approaches are easy to understand and use, they have limitations when it comes to controllability, efficiency, and speed. Neural network-based IST has become an important focus in style transfer research due to the development and application of deep learning. Without requiring creators to acquire specific skills or experience, this technology enhances production efficiency while facilitating the direct creation of genuine artistic styles. picture encoding is currently being used by several researchers for difficult multi-threaded operations in intelligent generation projects, such as picture synthesis, video synthesis, and text-to-image synthesis (Jin and Yang, 2021).

The author have successfully used style transfer algorithms to create lacquerware-themed creative products by compiling a set of lacquerware-related techniques, running simulations, testing demos, and adding secondary overlay effects using graphic software. Took advantage of two modules from generative adversarial networks (GANs) to create chair images (Liu et al., 2022). For this particular assignment, we used the first GAN to generate images of chairs and the second GAN to improve their resolution. Unfortunately, the collected chair photographs were shot from a variety of angles, thus the final idea images were missing some finer features. Have made comparable contributions to this area. We suggest a way to quickly create product renderings by combining Style GAN with drawing techniques (Cao et al., 2023). On the other hand, this method uses images that have been warped and sketches of the products to make exact models of the goods at different levels. Then, a GAN is used to take traits from these pictures and use them to make new design plans. If they do more work backwards, they might be able to turn the doodles into good colour designs (Nordin et al., 2022).

Pictures can have their styles transferred in two ways: randomly or by using a specific style. Training feedforward neural networks through model optimisation enhanced transfer speed but could only handle transfers of a single style. Subsequently, a plethora of academics put forth GAN-based visual style transfer approaches, with promising results. A single-style transfer method is what the aforementioned approach is.

3 Proposed method

Using deep learning, Figure 1 lays out a method for recognising and transferring artistic styles. Collecting and preparing data entails obtaining and standardising art datasets as the initial stage. Next, deep visual features are extracted using CNNs such as Res Net or VGG. The purpose of feeding these features into an artistic style recognition model is to classify artistic styles. Training and optimisation, which comprises tweaking hyper parameters and loss functions, subsequently enhance the model. Simultaneously (Li et al., 2023), a neural style transfer network optimises both the content and the style of the images, leading to stylised results. Accuracy, SSIM, and LPIPS are some of the metrics used to quantify the output in performance evaluation. Finally, the data is validated by results analysis, which involves statistical comparisons and investigations.

3.1 Deep visual feature extraction (CNN)

In order to maintain authenticity while still being creative, we remove the globally applicable style that is semantically valuable, excluding user interaction and low-level components. Then, we recover the colour-related data. In order to reduce computational complexity, the suggested method generates images based on full resolution. The level of professionalism will be raised by making use of a style transfer approach that has been employed before. Reducing the number of identical images was accomplished using a neural related style transfer method. Utilising VGG16 allows for efficient pretraining of convolutional neural networks. It is necessary to mix and create the various shades in one location in order to implement the style transfer in locations with different semantic contexts. Enhancing the gradient level image flow is the max-pooling technique. In a real-time context, an integral part of the creative process is utilised to produce art. The representations of style and content are implemented using the layers of VGG16, a pretrained neural network. Figure 2 shows the output of this neural method's use of

VGG-16, a network with 16 convolutional layers and 5 pooling layers, for the generation of artwork.

Figure 1 Methodological framework for artistic style recognition and style transfer (see online version for colours)

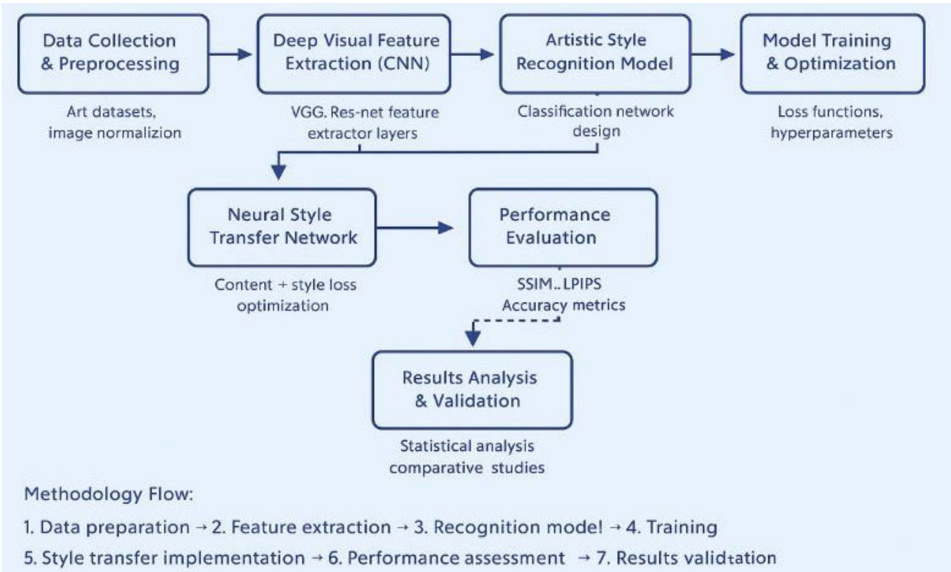
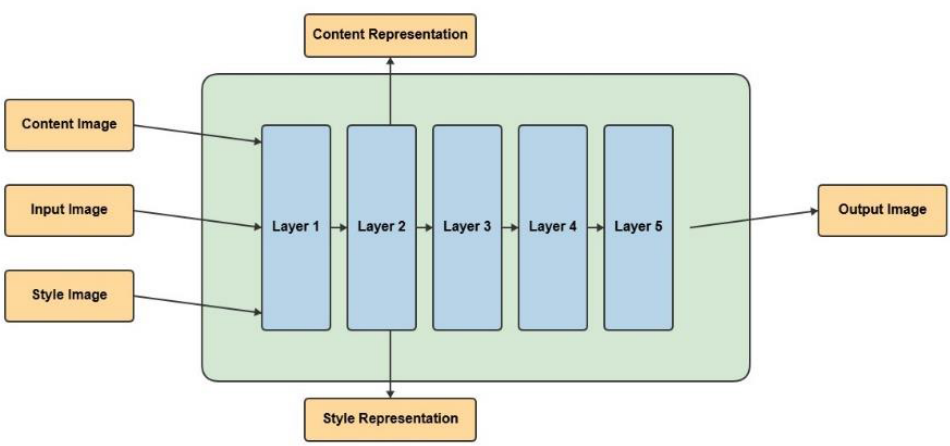


Figure 2 Block diagram of proposed system which takes input image (see online version for colours)



Both the input and style images are initialised to the content image whenever a photo is received via the network. The pixilated attributes, which stand for style, are extracted in the network's initial layers. During the extraction process, the layers extract the image's content. Once the style and content representations are recovered, the output is generated by reducing the losses between them. The final product is an image that is a combination of the input image's content and its style. Convolutional neural networks are useful for picture processing. In a layer, each neurone takes input, computes the dot product (Gușiță et al., 2023), and may or may not seek nonlinearity. By utilising neurones that are created in three dimensions – height, breadth, and depth – in its layers, CNN stands apart from all other neural networks. Activation volume, not network depth, is what 'depth' refers to here.

Using a feedforward technique, it consists of multiple layers of tiny computing units that produce visually unpredictable input. The main building blocks of a CNN are layers that use CNNs. Taking an input image and extracting a given feature is the job of these layers, which are collections of image filters. Their intended output is the feature map. CNNs with image processing training create an initial model of the input image based on object data and the image's processing capabilities; subsequently, they transform this model into pixel-based representations in order to produce high-quality images. By organising the higher-level contents and objects, the original image's pixel values are executed, and the lower levels are reconstructed using the input image's pixel values. The CNN brains are located in the convolution layer, which is responsible for merging datasets through mathematical processes. It takes input data and applies a convolution filter to create the feature map. Feature space is employed to limit the texture data in order to retrieve the image's style. The feature space is refined for each layer's filter responses. The correlations are contained inside the alternative filter representations.

The input image's texture components can be identified by integrating the feature correlations of various layers in its multiscale representation. To calculate the operation, the filter is slid over the input picture. A feature map is generated at that specific site by use of a component-related matrix construction. Figure 3 shows the typical filter in action, performing the common region of producing the convolution operation.

Figure 3 Images and filter (see online version for colours)

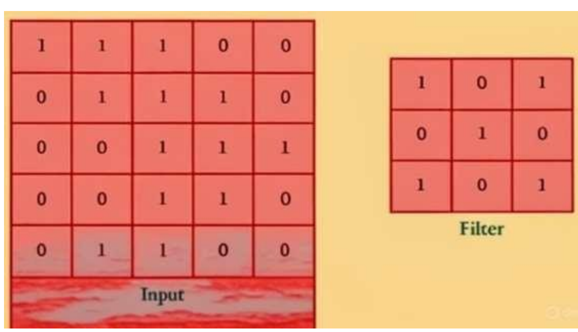
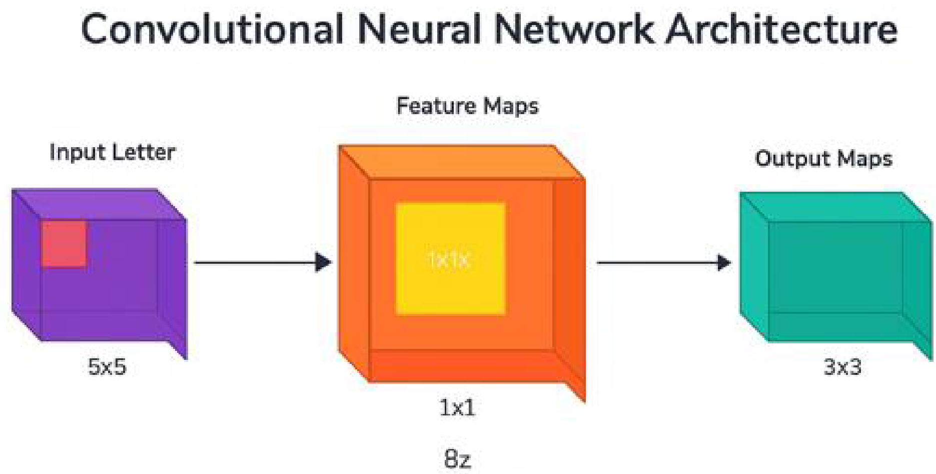


Figure 4 depicts a feature map that utilises convolutional input and filters. Filtering the aggregated input yields the feature map convolution results. The convolution processes are computed using three-dimensional (height, width, and depth) image matrices. A convolution filter may meet the needs of three dimensions with its specific width and height being 5 × 5 and 3 × 3. After the convolution layer generates an output, it uses several filters to create the feature maps that ultimately generate the result. The input must be determined in order for the convolution filter to function. Enclosing the input with zeros is done using the padding notion if equal dimension maintenance is required.

Figure 4 Input and filter convolution produces feature map (see online version for colours)



The CNN trains the restructuring loss ($Loss_R$) using the following equation: I is the input image. O is the target output picture.

$$Loss_R = \frac{1}{N}\sum_{i=1}^{N}\|O_i - I_i\|_2^2 \tag{1}$$

In the style branch, the perceptual loss $Loss_P$ is calculated using this formula:

$$Loss_P = \sum_l wW_l \|F_l(O) - F_l(I)\|_2^2 \tag{2}$$

$Loss_{st}$ stands for the restructuring loss for style, Sti for the stylised picture, $Loss_N$ for the normal image, and $Loss_{in}$ for intensity. Restructuring loss for style ($Loss_{st}$) can be calculated using this formula:

$$Loss_{st} = \sum_l w_l \|G_l(O) - G_l(S)\|_F^2 \tag{3}$$

The following equation is used to calculate the intensity-related restructuring loss:

$$Loss_{in} = \frac{1}{N}\sum_{i=1}^{N}|O_i - I_i| \tag{4}$$

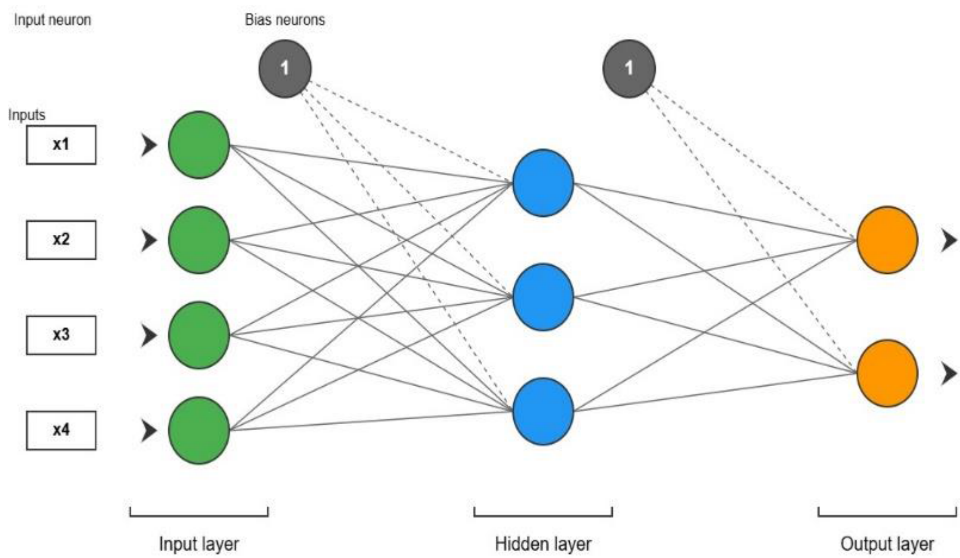
Networks that use nonlinear values to fix the position of their layers conduct complexity analyses. In each CNN layer, the input image is pre-arranged based on the image's filter reaction. The position of a layer is determined by mapping it with the different filters that produce feature maps for each matrix. An input image is used to obtain the style representation, and texture data is found by locating the feature space. A response filter in the layers has been built into the feature space. For feature maps that are spatially connected, it includes the anticipated correlations within the different filter answers. The following equation uses these correlations to frame the generalised matrix using a feature map:

$$G_l = F_l F_l^T \tag{5}$$

Artistic style recognition model

To begin, spots are used to separate the image. The next step is to put each patch into a category using a deep neural network. It's possible for every style class to happen. In the second step, the final style name for the whole piece of art is made using a shallow NN or a similar classifier (Imran et al., 2023). This way of naming styles is more accurate and useful because it looks at both the general style patterns and the more specific style traits in the art. If an artificial neural network only has a few layers between its input and output levels, it is known as a 'thin neural network'. Figure 5 show that most designs have an input layer, one or more hidden levels, and an exit layer. Many times, shallow NNs are used to do things like finding patterns, sorting things into groups, and regression (Liu et al., 2023). Deep neural networks (NNs) and CNNs are good at most machine learning tasks, but they might not be able to answer hard ones. They can help with some jobs since they are easier to understand and don't need as much work.

Figure 5 Shallow neural network fundamental design (see online version for colours)



Every deep level of a shallow NN has a lot of neurons. They can change the raw data in both simple and complicated ways because of their biases and weights. The network design, which includes how many hidden levels there are and how many neurons are in each layer, is usually chosen based on tests and knowledge of the subject. Weak NNs can handle more difficult tasks since they are made up of simple models. A simple neural network was chosen because it is easy to use and doesn't need a lot of computer power. You need less processing power to train and draw conclusions when shallow NNs are used. Deep CNNs were created to understand and organise visual data, mostly photos. Neural networks are being used in more and more computer vision tasks, like separating parts of a picture, recognising objects, and putting things into groups. You can change the things that deep CNNs are trained on. Some of these are the number of levels, the size of the screens, and the way the layers are arranged. A number of different ideas are used to build CNNs like Res Net, Inception, and VGG.

As a base for transfer learning, these networks are often used in many computer vision tasks because they were trained on large datasets. Deep neural networks (DCNNs) and shallow neural networks (SNNs) are both useful, but they are better or worse depending on the data and the situation.

3.2 Neural style transfer network

A few little structures from the original image remain in the stylised version; all they do is take on the hues and textures of the new image. Stylised versions of style images often include local structures that weren't there in the original, leading to an inaccurate portrayal of the style's artistic intent. The problem is that these techniques rely on feature extraction from high-resolution photos alone, without considering the content image when deciding what details to keep and what to remove. Our coarse network differs from other efforts in that it transfers low-resolution patterns in a crude form. This means there's more room to soak up low-frequency data, which is essential for figuring out the picture's general structure. When training is complete, some superfluous high-frequency data is disregarded. Figure 6 shows that when the coarse stylised image is transferred at a high resolution using the coarse network, more details that aren't necessary are transferred. The coarse network can eliminate certain irrelevant structural features and smooth up the items in the stylised image even at low resolution.

Figure 6 Vibrant landscape painting depicting a mountain range, forest, and lake under a vivid sunset sky (see online version for colours)

Note: The comparison was made for each image.

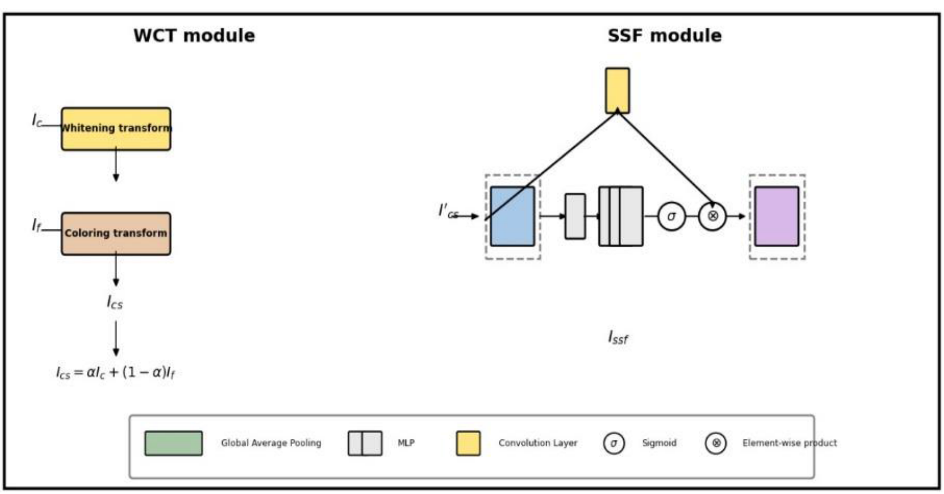
3.2.1 WCT module

We copy whitening and colouring transformation (WCT)'s low-resolution coarse-style pattern transfer with lightening and colouring changes in our network. If you clean something, it might get rid of style-related information that isn't needed while keeping the structure of the material. After the colour change, some style structures can be put together with content structures while still having the main visual style. Features in the lower layers only store data about colour and structure. When it comes to local patterns, higher-layer traits can record more complex ones. Because we change the colour and brightness of a single level to style, our coarse network is not the same as WCT. The styled parts aren't used to make the picture; they're used at different steps of putting it back together. By putting together rough features at different levels, our coarse network can save computer power while still gathering data at several levels.

3.2.2 Architecture of coarse network

Figure 7 shows the network as a whole, which is made up of encoders, WCT units, and decoders. A VGG- 19 network that has already been trained is used as the decoder during training. It puts out as a content feature and as a style feature at ReLU 4_1. It takes in and after that, a WCT tool is used to make the hair lighter and change its colour. Finally, a reconstruction decoder is used to bring back the rough artistic feature G". Another way to make the feature map as big as possible is to use the nearest neighbor upsampling layer in the same way that the VGG-19 network does. When _ is used as input, the rewriting method makes the updated styled features ( ). This rebuilding decoder gives you the first upsampling layer, the last convolution layer, and the upsampling layer that came before it. They are going to be fed into the fine network.

Figure 7 A WCT module and an SSF module are depicted in the schematics (see online version for colours)



4 Results and discussion

The tests were conducted using the MATLAB software, and the performance evaluation was implemented using the Wiki Art dataset. The style ambiguity is used to perform the style classification. In these studies, processing time, restructuring loss, and accuracy are utilised as performance indicators to compare the suggested methodology with analogous methods of Deeplab, CAN, and Seg EM. In comparison to comparable methods, the suggested methodology effectively improves accuracy, as shown in Figure 8, which shows the accuracy percentage for successfully completing style transfer.

Figure 8 Accuracy % (see online version for colours)

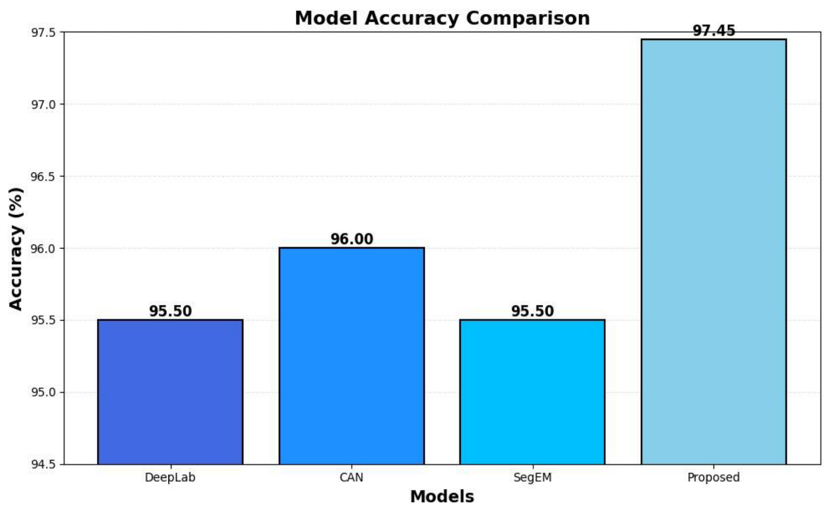
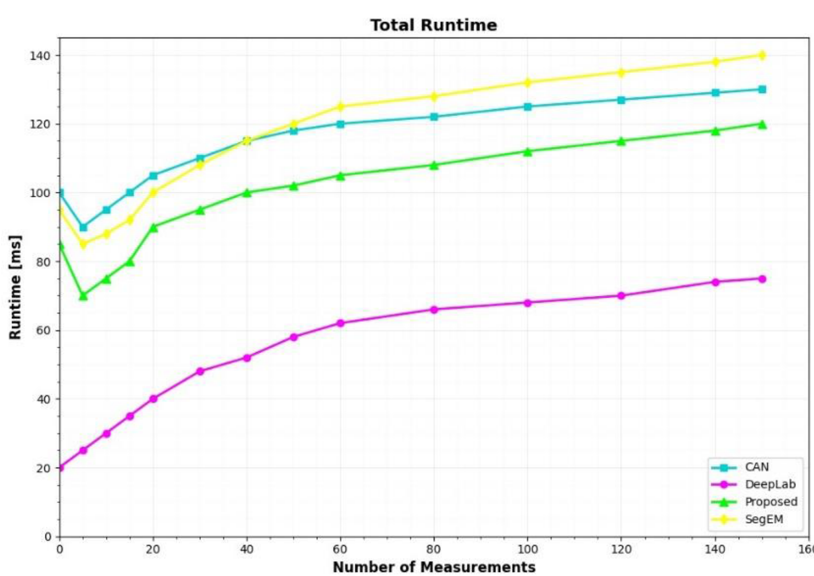


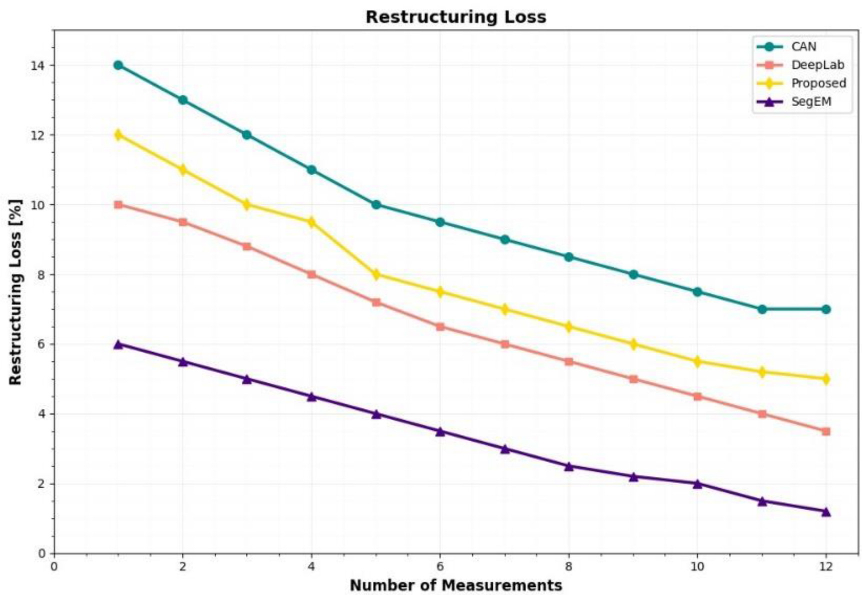
Figure 9 Total runtime (see online version for colours)



The results showed that the suggested methodology had a reduced amount of runtime compared to the other ways, and Figure 9 shows the runtime for the style transfer measurement. In order to make the suggested methodology more effective, the restructuring loss is a key metric.

Figure 10 shows that when compared to analogous approaches, the suggested approach results in a lower amount of restructuring loss %. The machine learning method's performance over time is shown by the curve, which is a graphic.

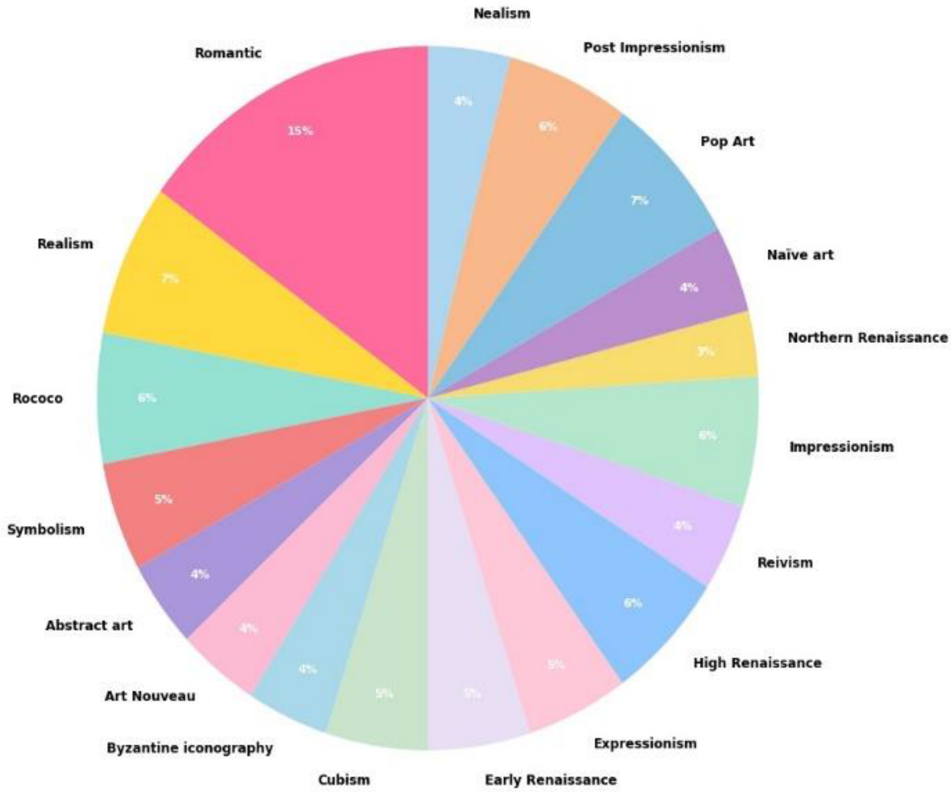
Figure 10 Restructuring loss % (see online version for colours)



The neural network based strategy has yielded remarkably good outcomes when using our suggested method. The proposed method uses massive amounts of data to create objects. The high-resolution, visually identical output image is the result of a properly completed style transfer. Doing away with style ambiguity loss is one approach to making the style transfer process more exact. We have proven that our proposed method can produce novel artefacts while preserving the quality of the images. These images served as a basis for categorising the various art trends. There are other ways to categorise the dataset's features. Two of these models include multiclass support vector machines (SVMs) and SNNs. Many art movements can be recognised and categorised using visual attributes utilising these methods. Collection 3 featured not just the classic Australian style but also images from the Pandora 18K collection. In all, 19,320 photos representing 19 different styles made up the dataset.

Figure 11 shows a very even distribution of photos throughout the various styles, visually representing the distribution of images. The Pandora 18K dataset has a lot of benefits, one of which was its high label validity. The Pandora 18K dataset differs from the Wiki Art collection in that art specialists assigned the labels rather than the general population. This improved the dataset's quality and dependability by making the labelling process more accurate and reliable.

Figure 11 Pandora 18K style dispersion percentages (see online version for colours)



Our approach takes cues from the EMD loss suggested by STROTSS and the whiting and colouring transformation offered by WCT, two recent methodologies that have been our inspiration. Figure 12 displays a comparison between our technique and WCT and STROTSS. While WCT is able to replicate the basic texture and colour distribution of images with arbitrary styles, it discards certain context local structure, leading to stylised images that are disorganised and unkempt (for example, rows 1, 2, and 3). However, the main colour scheme of the style picture is misrepresented, and there are too many structural elements kept (for example, rows 2 and 3). Our approach may transfer the fundamental structure while discarding some unimportant aspects of the content image, unlike these two alternatives. The stylised image's colour palette is indistinguishable from the style images, but the content image's global structure can borrow some obvious local features, like brushstrokes, from the style image.

For instance, the style pictures in the second and fourth rows are very similar to our stylised photographs, especially in terms of the mountain colour blocks and the flower brushstrokes. In order for our model to learn important style structures, we can ignore features about the content that aren't crucial.

Look at Figure 13 to see how we stacked up against other top-notch style transfer methods. Make the first method for transferring styles that is based on efficiency. It should be able to copy the style's look and how the colours are spread out. But some of the styled pictures (like rows 4, 5, and 6) have weird backgrounds that make them look wrong. They want to solve the problem in the same way that we do, which is through feed-forward. You can mix the local colours and designs of style pictures with the way the material is put together. On the other hand, it remembers too many structures and is only used to change the colour range in certain cases, like rows 1, 2, and 3. Two random style transmission models that mostly talk about simple style patterns are Ada IN and SA Net. Because of this, rows 4, 5, and 6 of SA Net are disorganised and have a rough surface. Also, Ada IN doesn't always show how the colours of style pictures are spread out. All of these methods keep a few small, unnecessary local structures of the content images, and the target picture doesn't have the style image local structures that it needs.

Figure 12 Comparison of generated artistic style transfer results across diverse scenes and structures demonstrating consistent preservation of content and stylistic features (see online version for colours)

Our model transfers the style colour distribution while merging local and global style structures, which has not been done before. For example, in the fourth row, our technology's stylised depiction of the rabbits looks more balanced and organic. Our approach can display the same creative expression as what appears to be an ink dot style image.

Figure 13 Illustration grid showing a series of stylized images transitioning through different colour tones and visual themes (see online version for colours)

5 Conclusions

Using deep visual feature extraction techniques, this research presented a new way for recognising artistic styles and transferring those styles to images. The suggested technique proved an efficient two-stage categorisation method for fine-art styles by combining shallow and deep neural networks. A WCT module was utilised to enhance the quality of style transfer while avoiding the loss of structural detail. Robust feature extraction was achieved through the usage of convolutional neural networks based on VGG. Results from the experiments validated that the proposed model achieved better results than the state-of-the-art methods with regard to processing time, restructuring loss, and classification accuracy. The model's scalability and dependability were further proven by using datasets with excellent label validity, such as Wiki Art and Pandora 18K. The model's capability to produce highly detailed styled images faithfully representing the desired style while meticulously maintaining the information is evidence of its practical application in creative image processing. In sum, the results of this research improve the capacity of machine learning to identify and transfer artistic styles, which in turn opens up exciting new avenues for digital innovation, automated art classification, and the protection of cultural assets. Investigating the model's potential use in interactive creative design tools and enhancing it to accommodate real-time style transfer are two potential directions for future research.

Declarations

All authors declare that they have no conflicts of interest.

References

Benzon, H-H., Chen, X., Belcher, L., Castro, O., Branner, K. and Smit, J. (2022) 'An operational image-based digital twin for large-scale structures', *Applied Sciences*, Vol. 12, No. 7, p.3216, https://doi.org/10.3390/app12073216.

Bica, I., Jarrett, D. and van der Schaar, M. (2023) *Invariant Causal Imitation Learning for Generalizable Policies*, arXiv 2023, arXiv:2311.01489.

Cao, M. et al. (2023) 'Recognition of occluded goods under prior inference based on generative adversarial network', *Sensors*, Vol. 23, No. 6, p.3355.

Ding, Y., Ma, X. and Yang, B. (2024) 'Research on image feature extraction and environment inference based on invariant learning', *Applied Sciences*, Vol. 14, No. 23, p.10770.

Gușiță, B. et al. (2023) 'Applications of the MPSA algorithm and CNN for securing medical X-rays', *2023 27th International Conference on System Theory, Control and Computing (ICSTCC)*, IEEE.

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.

Imran, S., Naqvi, R.A., Sajid, M., Malik, T.S., Ullah, S., Moqurrab, S.A. and Yon, D.K. (2023) 'Artistic style recognition: combining deep and shallow neural networks for painting classification', *Mathematics*, Vol. 11, No. 22, p.4564.

Jin, H. and Yang, J. (2021) 'Using computer-aided design software in teaching environmental art design', *Computer-Aided Design and Applications*, Vol. 19, No. S1, pp.173–183, https://doi.org/10.14733/cadaps.2022.S1.173-183.

Li, H., Huang, C. and Gu, L. (2021) 'Image pattern recognition in identification of financial bills risk management', *Neural Computing & Applications*, Vol. 33, No. 3, pp.867–876, DOI: 10.1007/s00521-020-05261-3.

Li, K., Yang, D. and Ma, Y. (2023) 'Image style transfer based on dynamic convolutional manifold alignment of halo attention', *Electronics*, Vol. 12, No. 8, p.1881.

Li, M. et al. (2020) 'Research on image style transfer technology based on semantic segmentation', *Comput. Eng. Appl.*, pp.622–627, DOI: 10.1109/CAC63892.2024.10865483.

Li, Z., Liu, F., Yang, W., Peng, S. and Zhou, J. (2022) 'A survey of convolutional neural networks: analysis, applications, and prospects', *IEEE Trans. Neural Netw. Learn. Syst.*, Vol. 33, No. 12, pp.6999–7019.

Liu, K., Yuan, G., Wu, H. and Qian, W. (2023) 'Coarse-to-fine structure-aware artistic style transfer', *Applied Sciences*, Vol. 13, No. 2, p.952.

Liu, L., Cui, J., Huan, Y., Zou, Z., Hu, X. and Zheng, L. (2022) 'A design of smart unmanned vending machine for new retail based on binocular camera and machine vision', *IEEE Consum. Electron. Mag.*, Vol. 11, No. 4, pp.21–31.

Ma, J., Yu, W., Chen, C., Liang, P., Guo, X. and Jiang, J. (2020) 'Pan-GAN: An unsupervised pan-sharpening method for remote sensing image fusion', *Inf. Fusion*, Vol. 62, No. 10, pp.110–120.

Nordin, H., Razak, B-A., Mokhtar, N. and Jamaludin, M-F. (2022) 'Feature extraction of mold defects on fine arts painting using derivative oriented thresholding', *Journal of Robotics, Networking and Artificial Life*, Vol. 9, No. 2, pp.192–201, https://doi.org/10.57417/jrnal.9.2_192.

Ranftl, R., Lasinger, K., Hafner, D., Schindler, K. and Koltun, V. (2020) 'Towards robust monocular depth estimation: mixing datasets for zero-shot cross-dataset transfer', *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 44, No. 3, pp.1623–1637.

Wang, X., Wang, W., Cao, Y., Shen, C. and Huang, T. (2023) 'Images speak in images: a generalist painter for in-context visual learning', *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Vancouver, BC, Canada, 17–24 June, pp.6830–6839.

Wang, X., Yang, L.T., Song, L., Wang, H., Ren, L. and Deen, M.J. (2021) 'A tensor-based multiattributes visual feature recognition method for industrial intelligence', *IEEE Transactions on Industrial Informatics*, Vol. 17, No. 3, pp.2231–2241, DOI: 10.1109/tii.2020.2999901.

Yunfei, D. and Li, B. (2020) 'Lightweight feature fusion network design for local feature recognition of non- cooperative target', *Infrared and Laser Engineering*, Vol. 49, No. 7, DOI: 10.3788/irla.2020-0170.20200170.

Zhao, Y., Samuel, R-D-J. and Manickam, A. (2022) 'Research on the application of computer image processing technology in painting creation', *Journal of Interconnection Networks*, Vol. 22, Supp05, p.2147020.

Zhou, Y. et al. (2024) 'Style transfer of Chinese Wuhu iron paintings using hierarchical visual transformer', *Sensors*, Vol. 24, No, 24, p.8103.