Automatic Ink Painting Rendering Technique Based on Deep Triangular Convolutional Network

Fu Wang[1,2*]
[1]Xuchang Vocational Technical College, Xuchang 461000, China
[2]Henan Network Engineering Intelligent Service Engineering Research Center, Xuchang 461000, China
E-mail: wangfu86@126.com
[*]Corresponding author

Keywords: automatic rendering, deep convolutional neural networks, digital images, feature extraction, ink painting

Received: September 9, 2024

In order to generate works with rich details and unique ink style in a short period of time, a deep triangular convolution network based on deep triangle convolution is proposed. Firstly, convolutional neural networks are used to extract feature textures from stylized images and apply them to noisy images to generate ink painting textures with specific styles. Then, by constructing a weighted sum of content loss and style loss, the effect of automatically adding ink painting style to the content image can be achieved. The deep triangular convolutional network structure adopted can effectively avoid the problem of gradient vanishing and improve the training efficiency of the network. The experimental results show that the testing time of the deep triangular convolutional network is 1.5 seconds, and the classification error rate is only 3.5%. The accuracy of the network reached 95.5%, significantly higher than Style Migration and others. The proposed method maintains high recognition accuracy and image quality while achieving style conversion in ink painting. The research provides new directions for the development of art, which helps to inherit and disseminate Chinese ink painting art.

Povzetek: Predlagana je tehnika avtomatskega upodabljanja s pomočjo globokega trikotnega konvolucijskega omrežja (DTCN). Metoda izboljša prepoznavanje umetniških značilnosti in omogoča kakovostno digitalno obdelavo stiliziranih slik.

1 Introduction

With the vigorous development of information technology, deep learning for artistic style images has become an integral part of machine learning. However, the current learning of artistic images is limited to the mechanical transfer of pixel features, and there has been limited in-depth research in the field of artistic style itself. Both artistic deep learning and technical operations are still in their early stages of development [1]. In the field of image processing, machine learning and neural network algorithms have emerged as promising methods, and they have gained significant attention from researchers [2]. Neural networks are computational models that mimic the information processing and learning mechanisms of the human nervous system. They consist of a large number of artificial neurons and their interconnections, and have been extensively studied and applied in classification, prediction, and data processing domains [3]. Artificial intelligence assistants have gradually become an effective tool to assist artists in the traditional process of ink painting creation, which takes a long time, involves a lot of manual intervention, and the quality of painting is limited by the artist's personal skill level. Applying deep learning technology to ink painting creation aims to achieve automated, efficient, and high-quality ink painting, which has important research significance and practical value.

In this study, through in-depth analysis of the characteristics of traditional ink painting, a new optimization network based on deep convolutional neural network (CNN) is proposed, called deep triangular convolutional network (DTCN). This model is used to extract artistic features from ink wash paintings, and then fuse the extracted features with human and landscape images, ultimately achieving the rendering application of digital image ink style while obtaining high-quality ink wash images. Through experimental verification, the proposed method has achieved significant results in terms of drawing quality, drawing speed, and automation level. In addition, this technology can also be extended and applied to other art fields, including automated creation of traditional Chinese painting, animation production, etc., providing new technical support for the development of the digital art industry.

2 Related works

Manual image discernment and processing can be time-consuming. However, since the advent of neural network algorithms, there have been significant research efforts to automate image processing, thereby freeing up human resources and improving the accuracy and speed of image processing. A W Q and their team proposed an image-based bubble tracking algorithm for plate heat exchangers, utilizing CNN to process images and obtain the position and state of bubbles in the flow. Based on this information, the algorithm calculated the movement and velocity of the bubbles. Even in scenarios with a large number of bubbles and dense distributions, this method achieved an average accuracy exceeding 94% and a recall rate exceeding 87% [4]. Peng Y et al. introduced a gated CNN pre-training algorithm for high-resolution reconstruction of remote sensing images. The algorithm consisted of several residual blocks with skip connections. Each residual block contained a specially designed gated convolutional unit that assigns different weights to high-frequency and low-frequency information, controlling information transmission and allowing the main network to focus on learning high-frequency information. Experimental results on various remote sensing datasets demonstrated the accuracy and visual improvement of the proposed PGCNN, outperforming several state-of-the-art methods [5]. Thangarajan S K and collaborators proposed a hybrid classifier based on CNN for assisting in the diagnosis of brain tumors using magnetic resonance imaging. In this algorithm, CNN utilized a three-level discrete wavelet transform input with optimized hidden neurons. Experimental results showed that the proposed algorithm achieved a 17% higher accuracy compared to the Support Vector Machine (SVM) algorithm and a 15.5% higher accuracy compared to traditional CNN [6]. OD A and their team addressed the diagnosis of glaucoma by proposing a diagnostic method based on graphs and CNN. CNN played a crucial role in interpreting enhanced image data. In the ORIGA dataset, the proposed method achieved accuracy, the area under the curve (AUC), and precision all exceeding 90%, with the accuracy reaching 82% in practical diagnosis [7]. Ghosh S K and colleagues introduced a novel retinal image segmentation method using a CNN based on a Ranked Support Vector Machine (rSVM) to detect diabetic retinopathy. The design of CNN incorporated rSVM to define consistent class labels within the network, reducing the number of channels that aid in rapid convergence. With the help of post-processing steps, the method achieved good segmentation accuracy exceeding 90% in different databases [8]. Deshpande A M and their team constructed a deep neural network based on global residual learning to automatically remove various types of stripes and grain noise from remote sensing images, thereby enhancing their quality. The proposed method exhibited improved performance compared to existing techniques in terms of peak signal-to-noise ratio and structural similarity index measurements [9].

There are also certain research achievements in the field of utilizing intelligent algorithms to process art works. Xie X and the partner developed a painting art style rendering system based on generative adversarial networks. This algorithm can suppress the limitation of paired samples and improve the quality, stability, and convergence speed of images generated by painting art styles [10]. Platkevi A led his team to propose an example-based animation style rendering method. This method used moving examples to extract dynamic features of animation and achieve natural transitions between single frames [11]. In order to accurately extract the artistic style features of art works, W Yao and the partner developed a color feature-based color extraction method. This approach combined corner detection, edge contour feature detection methods, and mainly used the fuzzy clustering method for style extraction [12]. The summary table for related works is shown in Table 1.

Table 1: Summary table for related works.

Researchers	Method	Results	Limitation
Wang et al [4]	Image based bubble tracking algorithm for plate heat exchanger	The average accuracy exceeds 94%, and the recall rate exceeds 87%	Affected by factors such as image quality and bubble morphology
Peng [5]	Gate controlled CNN pre training algorithm	Image quality improved by 11%	Applicability on different scenarios and datasets
Thangarajan [6]	Hybrid classifier based on CNN	Compared to traditional CNN, the accuracy has increased by 15.5%	Affected by factors such as image quality and tumor type
Deperlioglu [7]	A glaucoma diagnosis method based on graphics and CNN	Accuracy, AUC, and precision all exceed 90%	Limited sample and clinical data
Ghosh [8]	CNN based on rSVM	The segmentation accuracy exceeds 90%	Affected by factors such as image quality and lesion morphology
Deshpande [9]	Deep neural network based on global residual learning	The structural similarity index can reach 85%	Not suitable for complex scenes and noise types
Xie [10]	A painting art style rendering system based on generative adversarial networks	The image quality can reach 90%, and the stability can reach 92%	Sample and artistic style limitations
Platkev [11]	Example based animation style rendering method	Realize natural transitions between single frames	Example quality and animation style
Yao [12]	Art style feature extraction method based on color features	The accuracy can reach 86%	Applicability of different art styles

Analyzing the research conducted in recent years on image processing using neural networks, most of these studies have focused on remote sensing image processing and diagnostic applications in medical imaging. The application of CNN-based image processing techniques in the field of art is relatively less explored. The conclusions drawn from industrial and medical fields cannot be directly extrapolated to the domain of art. In the field of art, there is also a certain gap in the application of intelligent algorithms that highly focus on specific artistic styles. Therefore, in this study, addressing the automatic rendering of ink painting images, a method based on deep CNN is proposed to fill this research gap in the art domain. The aim is to lay the groundwork for practical research outcomes in this area.

3 Deep CNN and its optimization

3.1 CNN algorithm

CNN achieves the requirement of minimizing data preprocessing by exploiting the connections between object features and reducing the number of parameters that need to be learned. CNN consists primarily of convolutional layers, pooling layers, and fully connected layers, as shown in Figure 1 [13]. In the hierarchical structure of CNN, image information as the input at the lowest level is processed in a specific order or propagated to different layers of the network [14]. The most salient features of the observed data in each layer can be extracted using filters. The schematic diagram of the model is shown below:

As shown in Figure 1, the process of transferring object data to different network layers follow a specific order. The primary step in the data transmission process is sending the original image to the input layer [15]. This is followed by combining the local features extracted from the local receptive field to generate feature maps in the C1 layer. Next, there is a summation, weighting, and function activation process to generate feature maps in the S2 layer. The subsequent steps continue the specific flow to realize the C3 layer and S4 layer. The final output data is obtained by converting the output feature maps into vectors through the fully connected layer. Specifically, the calculation Equation for the convolutional layer in CNN is as follows:

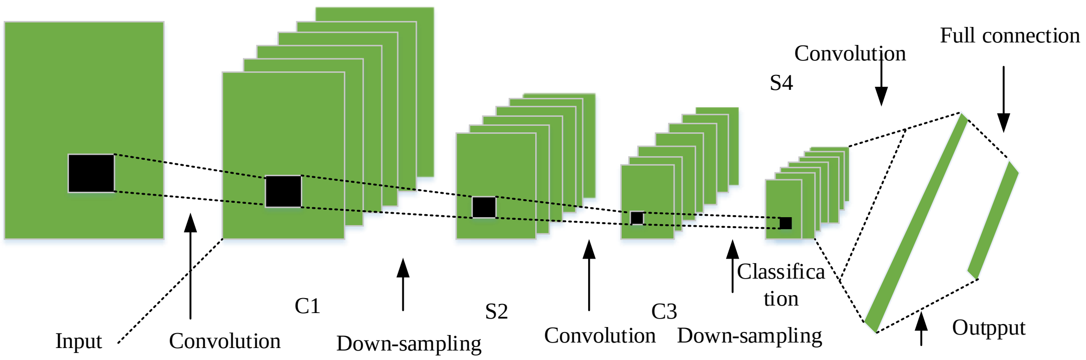


Figure 1: Deep convolutional neural network.

$$X_j^l = f\left(\sum_{i\in M_j} x_i^{l-1} * k_{ij}^l + b_j^l\right) \quad (1)$$

Its algorithm learning consists of three levels: convolutional layer computation, pooling layer computation, and weight updates [16]. In the convolution process, the basic kernel convolves the original input image, and then the new feature maps are obtained through the activation function's non-linear mapping [17]. Different convolution kernels can extract different feature maps. In Equation 1, k_{ij}^l denotes the convolution kernel of layer l, l represents the convolutional layer number. X_j^l represents the j th output feature map in layer l, and the activation function is denoted by $f()$. The convolution operation is represented by *, and b_j^l is the number of feature maps in layer $l-1$. Additionally, the Equation for the pooling layer computation is as follows:

$$x_j^l = f\left(\beta_j^l down(x_j^{l-1}) + b_j^l\right) \quad (2)$$

Connected feature maps have corresponding relationships in the pooling layer [18]. To reduce the amount of data processing and retain useful information, feature maps need to be downsampled. In the Equation, $down()$ represents the downsampling function, and β represents the downsampling coefficient. Finally, the Equation for the weight update function is as follows:

$$W(t+1) = W(t) + \eta\delta(t)x(t) \quad (3)$$

After determining the residual calculation, further updates are required for weights and biases. Assuming that the current moment of CNN is t and the next moment is $t+1$, the Equation for weight update can be obtained. In the Equation, η represents the learning rate, $x(t)$ represents the input to the neuron, and $\delta(t)$ represents the error term. In summary, the algorithm learning of deep CNN is achieved through convolutional layer computations, pooling layer computations, and weight update computations. The algorithm learning directly impacts the training process of CNN.

3.2 Optimization

For a specific CNN, it should have an overall smooth downsampling [19]. In this study, an improved and optimized deep triangular CNN with triangular structural units is proposed based on the foundation and achievements of traditional convolutional networks. Its main advantage is that it can gradually increase the dimensionality of feature maps through neural layers in the network. Using a deep triangular network can maximize the propagation of feature information in the backward process and greatly improve the utilization of feature maps. In simple terms, the triangular network plays a significant role in avoiding the gradient vanishing phenomenon. The system structure diagram is shown below:

The objective of this model is to achieve the dispersion of structural unit pressure by increasing the number of feature maps. As shown in Figure 2, the entire DTCN model consists of six modules. In each module, the first pooling layer uses max pooling, while subsequent layers use average pooling. The dimension of the feature map in the module output channel shows a linear growth trend, similar to a triangle. This network achieves a gradual increase in the dimensionality of the feature map while maintaining the original convolution kernel size unchanged. This enables the network to more effectively capture complex features in images, thereby improving recognition accuracy. Secondly, the triangular structural units in deep triangular networks help to improve the stability of gradient vanishing phenomena, enabling the network to converge faster during training. The schematic diagram of the feature map dimensionality at specific layers is shown below.

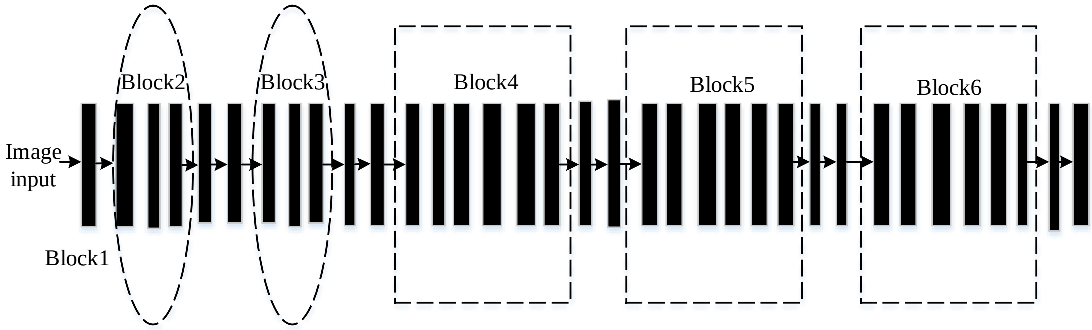


Figure 2: The deep triangular convolutional network structure.

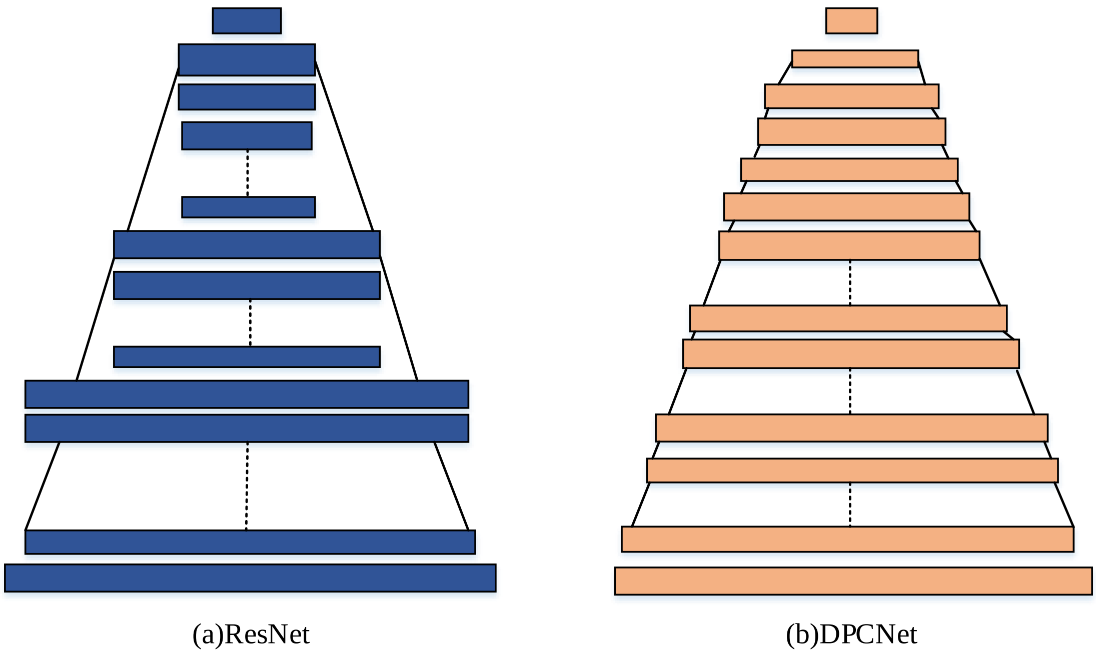


Figure 3: Schematic diagram of feature dimension.

Table 2: DTCN feature dimension.

Output feature map size	**32*32**		**16*16**		**8*8**	
Network layer	Block1_conv1	Block2_conv2	Block2_pool2	Block3_conv3	Block3_pool2	Block4_conv4
Design	3*3 conv,64	[3*3 conv,64+aN]	Max 2*2("same")	[3*3 conv,64+aN]	Ave 2*2("same")	[3*3 conv,64+aN]

Figure 3: Shows the schematic diagram of the feature map dimensionality for the triangular structure of ResNet and DTCN. It can be clearly seen that starting from the input layer, the feature map dimensionality in the residual network module remains unchanged, and this continues until the downsampling layer where the dimensionality of the feature maps doubles. ResNet and DTCN are both composed of CNNs, but DTCN's connections allow for direct information transfer across layers, making network training more efficient and easier. Currently, DTCN is being widely applied in computer vision tasks such as image classification, object detection, and semantic segmentation. As a result, DTCN is the main focus of this study. The feature map dimensionality structure of DTCN is shown in Table 2. As the network layers change, there are significant variations in the design and output feature map sizes.

In the image processing process of DTCN, it is necessary to provide the Equation for computing network feature maps. Assuming an image passes through a CNN, and the network has L+1 layers. Here, shortcut connections are used in a concatenation manner to obtain the network feature maps. The use of concatenation allows any layer in the network to access the information from previous layers, and its input can be forwarded to subsequent layers. Under these settings, the Equation for computing the feature maps in the L+1 layer of the network is given by equation (4).

$$\mathrm{x(L+1)}=\mathrm{H}_{L+1}([x(0),x(1),\cdots,x(L+1)]) \tag{4}$$

In the Equation, $\mathrm{x(L+1)}$ represents the feature map of the image x in the $\mathrm{L+1}$ th layer. $\mathrm{H}_{L+1}()$ represents the non-linear transformation operation. The image first enters the convolutional layer and generates feature maps through convolution operations. Then, pass these feature maps to the next layer, while using shortcut connections to combine the feature maps of the previous layer with the input of the current layer. In each layer, the feature maps undergo nonlinear transformation operations to generate new feature maps, which are then passed on to the next layer. In this way, the feature map propagates layer by layer in the network until it reaches the final output layer. Throughout the process, the dimensionality of the feature map increases linearly with the increase of network layers until the final output result is achieved. Additionally, the activation function is improved based on the ReLU function.

$$\mathrm{ReLU(x)=max(0,x)} \tag{5}$$

The ReLU activation function is primarily effective in eliminating non-negative elements. However, the use of activation functions also has its limitations, such as causing filters to generate a large amount of redundant information. As this can impact the accuracy of the research, this study introduces a new activation function based on ReLU to effectively eliminate pairing phenomena and various redundant information, meeting the requirements of DTCN. The mathematical expression of this new activation function is as follows:

$$\mathrm{CReLU}(x)=\left[\mathrm{ReLU}(x),\mathrm{ReLU}(-x)\right] \tag{6}$$

The main advantage or functionality of the new activation function proposed in this study is that it can reduce the information redundancy caused by non-linearity in filters and also preserve the positive and negative phase information extracted by the filters. In the final stage of neural network design, the dropout technique is used to prevent potential overfitting issues in CNN [20]. Dropout randomly deactivates some neurons in the neural network by setting their outputs to zero during the training process. Specifically, each neuron has a certain probability of being dropped or retained. This can reduce complex dependencies between neurons and enhance the network's generalization ability [21]. The randomness of dropout prevents the neural network from overly relying on specific neurons, forcing the network to learn multiple independent features, thus improving model robustness and generalization ability. Additionally, using dropout helps reduce the number of parameters in the network, making it more concise and efficient. When the Dropout layer is set below the identity mapping module, there is no significant change in the performance of the neural network. Therefore, the Dropout network is set between the two convolutional layers in the triangular structure unit, as shown in Figure 4.

Figure 4 shows the difference between DTCN without a Dropout layer and the network with a Dropout layer, where the Dropout layer is set between different convolutional layers.

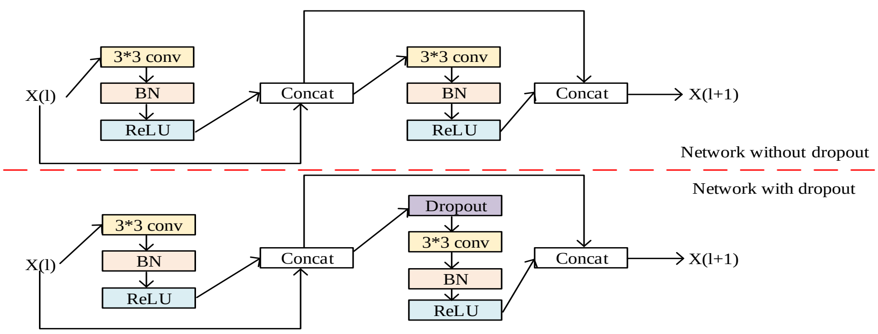


Figure 4: Triangular structure unit with Dropout.

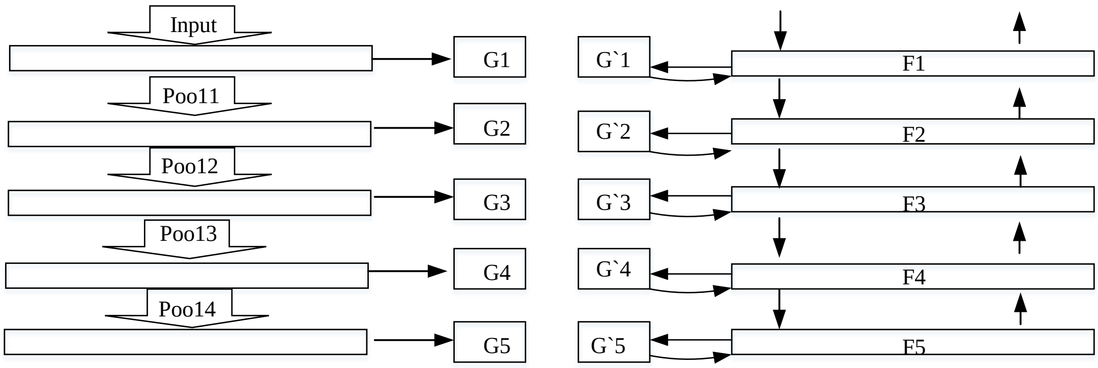


Figure 5: The process flowchart for generating feature textures.

4 Implementation of automatic ink painting rendering technique

4.1 Generation of feature texture

From a computer science perspective, artistic creation can be translated into the encoding process of applying specific algorithms to artistic content. Based on this idea, this study utilizes CNN to generate specific textures of artwork from noise images, achieving the process of feature texture generation. The process flowchart for generating feature textures based on deep CNN is shown in Figure 5 below.

The key steps for generating texture features from style images in the feature texture generation process can be clearly seen in the texture generation diagram. Firstly, the image is input into the VGG-16 network. Then, the crucial step is to calculate the feature responses F^1 for each layer, based on which the matrix G^1 representing the texture features of the style image is computed. Following the same principle or rule, the primary step in generating texture features from noise images is to calculate the feature responses $\hat{F}^1$ using VGG-19, resulting in the matrix $\hat{G}^1$. In the process of generating texture features from style images, E^1 represents the variance and loss of each layer in the network, and the image pixel $\hat{x}$ is updated through gradient descent. The matrix correlation Equation is as follows:

$$G = \begin{bmatrix} g_{11} & \mathrm{g}_{12} & \cdots & \mathrm{g}_{1\mathrm{n}} \\ g_{21} & \mathrm{g}_{22} & \cdots & \mathrm{g}_{2\mathrm{n}} \\ \vdots & \vdots & \ddots & \cdots \\ g_{n1} & \mathrm{g}_{\mathrm{n}2} & \cdots & \mathrm{g}_{\mathrm{nn}} \end{bmatrix} \tag{7}$$

After obtaining the feature planes of the image from deep CNN, the next step is to consider how to utilize these features to describe the style texture of the image. The basic method can be summarized as building feature layers to calculate the correlations between different filter responses. In other words, matrix correlation is required in the actual texture generation process of the style image. For ease of calculation, the results can be normalized using the following Equation.

$$E = \frac{1}{4N^2M^2}\sum_{i,j}(G_{ij} - A_{ij})^2 \tag{8}$$

Once E is obtained, the following operations can be performed to complete the normalization.

$$Lstyle = \sum_{l=0}^{L} w_l E_l \tag{9}$$

By optimizing the algorithm, an image with the same style texture features as the original image can be generated. From equation (7), it is clear that the original image is denoted as A, while the noise image is represented by G. N denotes the number of feature maps, and M represents the size of the feature maps. The loss value of w_l serves as a weighting factor for the overall loss value, and a smaller value of w_l can weaken the influence of that layer.

4.2 Ink painting style rendering

Based on CNN, style rendering should follow specific principles that consider both user aspects and system development and deployment. Therefore, for ordinary users, this study aims to ensure good stability and convenience. The rendering module, as the core module in this research, mainly achieves the rendering functionality for ink painting art style. The specific structure diagram of the style rendering is in Figure 6.

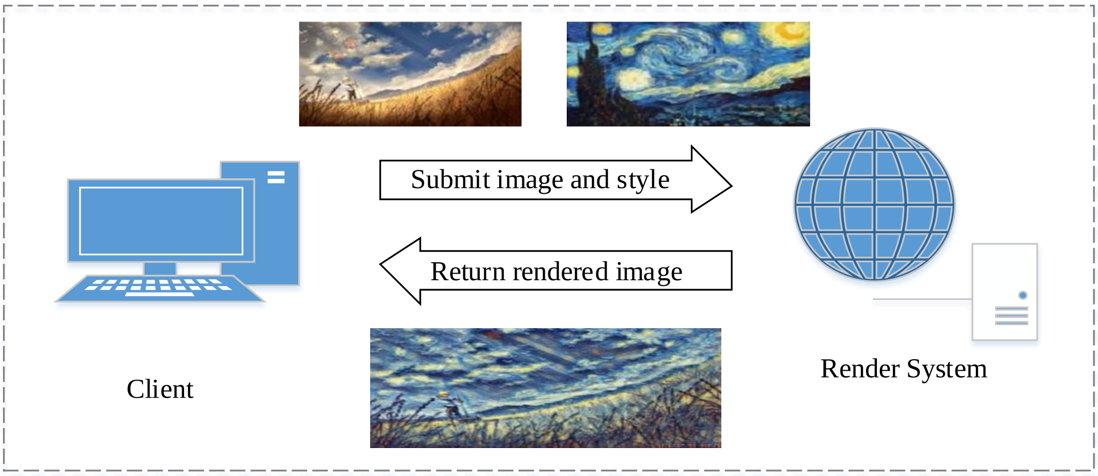


Figure 6: Style rendering structure.

After completing the generation of ink painting texture, it is necessary to render arbitrary images and generate ink painting style texture automatically. The main goal of this research is to achieve ink painting style effects on content images through rendering, and the basic approach or process to achieve this is to determine a total loss, which is the sum of content loss and style loss. Minimizing the total loss involves minimizing the sum of the noise image with the content image and the ink painting image. The mean squared error loss Equation for defining the content image is as follows:

$$Lcontent = \frac{1}{2}\sum_{ij}(F_{ij} - P_{ij})^2 \quad (8)$$

In the Equation, the content image is denoted as p. The response of filter i at position j in the content image and the noise image are represented as P_{ij} and F_{ij} , respectively. It should be noted that when the colors and thickness variations in the generated ink painting image appear abrupt or disharmonious, the total loss can be calculated by computing the mean squared difference between neighboring pixels. The calculation Equation is as follows.

$$L_{total} = \alpha L_{content} + \beta L_{style} + \gamma L_{noise} \quad (9)$$

In the Equation, L_{total} represents the total loss, $L_{content}$ represents the content loss, L_{style} represents the style loss, L_{noise} represents the noise loss, and α, β, γ represent the weighting factor of the content image, style image, and noise image.

5 Analysis

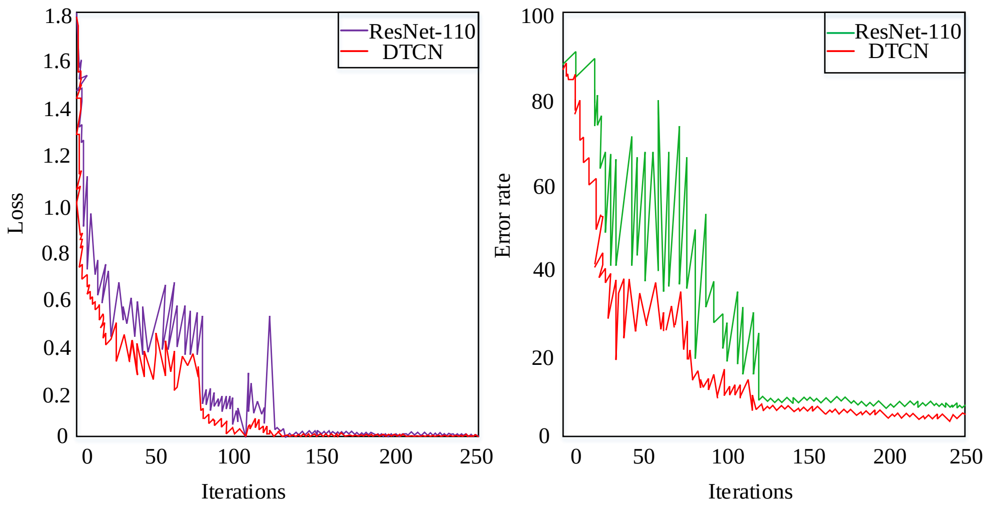
5.1 Comparison of residual network and DTCN performance.

Research shows that although residual modules can achieve or train deep networks based on their unique capabilities, this does not mean that residual networks are without drawbacks when it comes to output data. On the contrary, compared to plain networks, residual networks have inherent limitations and weaknesses. The main reason for this is that in residual modules, gradients are not forced to pass through the weights of the residual module while flowing through the network, making it difficult to achieve deep learning of useful information during the training process. The experimental analysis and results of this research are divided into two parts: the analysis of the proposed triangle network and the application effect of the proposed network in automatic ink painting style rendering. Firstly, the software and hardware environment for the experiments are specified, as shown in Table 3. To ensure smooth operation and prevent hardware limitations from affecting algorithm performance, the experimental machine is equipped with 32GB of memory. Since the experiment involves image processing and rendering, an NVIDIA GTX 3060TI GPU is used to ensure rendering efficiency.

The proposed triangle network in this research provides a solution to address these problems, and its performance is compared in Figure 7. From the specific training loss function curve and classification error rate curve in Figure 7, it can be observed that the triangle network has a lower training loss function value compared to the residual network. The comparison of the training loss functions between the two networks indicates that the triangle network has smaller differences between the ground truth and predicted values. This validates that the triangle network structure possesses stronger network optimization capabilities, its features can be effectively and sufficiently utilized, and the network converges faster. Additionally, the triangle network structure exhibits lower classification error rates compared to the residual structure, indicating better generalization ability.

Table 3: Experimental environment.

Experimental environment		Information
Hardware	CPU	Intel 12400
	RAM	32 GB
	GPU	GTX 3060TI
Operational system		Windows 10
Development tool		Java
Analysis tool		Matlab



(a) Training loss (b) Error rate of classification

Figure 7: Performance comparison between residual networks and deep triangular convolutional network.

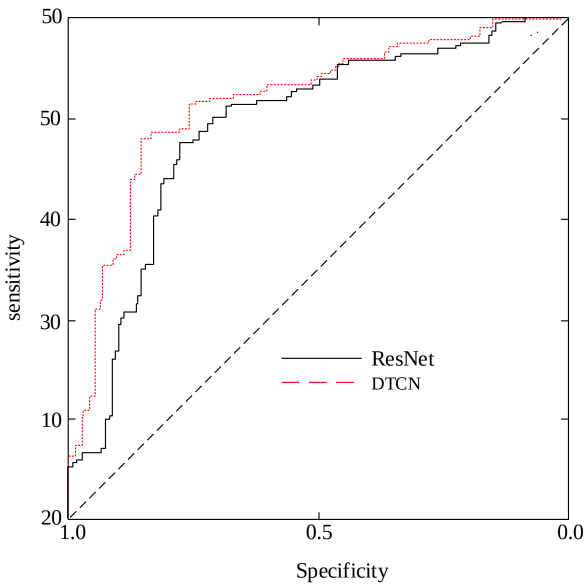


Figure 8: Receiver operating characteristic of the networks.

After analyzing and comparing the loss function curves of the algorithm, the performance and effectiveness of the algorithm are further evaluated by analyzing the Receiver Operating Characteristic (ROC) curves. ROC curves illustrate the classification accuracy of the model at different thresholds. ROC curves provide an intuitive representation of the overall performance of the classifier at different threshold settings and serve as a benchmark to select the optimal threshold. The ROC curves of the algorithm and its comparison with ResNet are shown in Figure 8. By examining the ROC curves of the two networks, it can be observed that the curve of DTCN is consistently above the curve of ResNet, and the AUC or DTCN reaches 0.801. AUC is a metric used to quantify the accuracy of model classification.

During the final stage of testing the DTCN network, its image processing capabilities on a real image dataset are analyzed, and the results are presented in Table 4. The dataset introduces for the experiment here is the Public Figures Face Database, which is a public data image from Columbia University with a data volume of approximately 58000 images. During the experiment, the segmentation ratio of the image dataset is 7:3. 70% of them are training sets, and the rest are testing sets. To ensure comparability of the results, the dimension of the input feature maps is set to 256 for each test. Firstly, normalize the image by scaling the pixel values to a range of 0 to 1 to eliminate the influence of brightness differences between different images; Next, contrast enhancement, sharpening, and other techniques are used to improve image clarity and further optimize recognition performance. With a parameter count of 30.1, DTCN exhibited the shortest testing time and the lowest classification error rate. At this point, the testing time is 1.5 seconds, and the classification error rate is only 3.5%. Overall, based on the information in the table, the classification error rate of DTCN does not exceed 5%, with a maximum of 4.9%. The testing time also remains within 60 seconds. In practical applications, a theoretical error rate below 5% is generally considered sufficient to complete most image rendering tasks.

Table 4: DTCN performance under different parameter count.

Parameter count(M)	**30.1**	**54.5**	**92**	**136.7**
Depth	18	24	30	36
Training time (s)	25.9	34.7	45.5	59.3
Testing time (s)	1.5	2.1	2.5	3.3
Input feature map dimension	256	256	256	256
Output feature map dimension	512	608	703	800
Classification error rate (%)	3.5	4.6	4.8	4.9

In terms of computational complexity, DTCN has significant advantages over standard CNN and GAN. Compared to standard CNN and GAN, which often have millions of parameters, DTCN has significantly smaller parameter sizes. This enables DTCN to converge faster during training and inference, reducing the demand for computing resources. DTCN has a lower FLOP and requires less computation during training and inference. Compared to standard CNN and GAN, DTCN has particularly obvious advantages in FLOP. This makes DTCN more efficient in practical applications and can complete image rendering tasks faster. By directly comparing with standard CNN and GAN, it can be found that DTCN significantly reduces computational complexity and parameter count while maintaining a low classification error rate, thereby improving overall efficiency.

5.2 Convergence of the loss in automatic ink painting rendering

The main goal of this research is to achieve ink painting style effects on content images through rendering, and the basic approach or process to achieve this is to determine a total loss, which is the sum of content loss and style loss. Minimizing the total loss involves minimizing the sum of the noise image with the content image and the ink painting image. The α/β ratio determines the similarity between the noise image and the content image, where a larger ratio indicates a closer resemblance. Conversely, a smaller α/β ratio leads to a higher degree of rendering. This research visualizes the ink painting rendering images and their convergence of loss processes for α/β ratios of 1/30, 1/50, and 1/200.

As shown in Figure 9, the visualized ink painting rendering images and their convergence of loss processes are analyzed for α/β ratios of 1/30, 1/50, and 1/200. The final content loss value is less than 1200e+6, the style loss fluctuates around 2000e+6, the noise loss fluctuates around 3000e+4, and the total image processing loss fluctuates around 4000e+6. Specifically, compared to the style image, the noise image exhibits a larger range of variation in the loss, with rapid increases and decreases, ultimately converging gradually. On the other hand, the style image loss and the overall loss demonstrate fast convergence characteristics. In general, the loss values for image processing at various α/β ratios converge to a low level. Typically, the quality of the rendered images can be evaluated by the magnitude of the loss, where smaller losses indicate higher quality rendered images. Additionally, the ink fusion loss function is defined as follows.

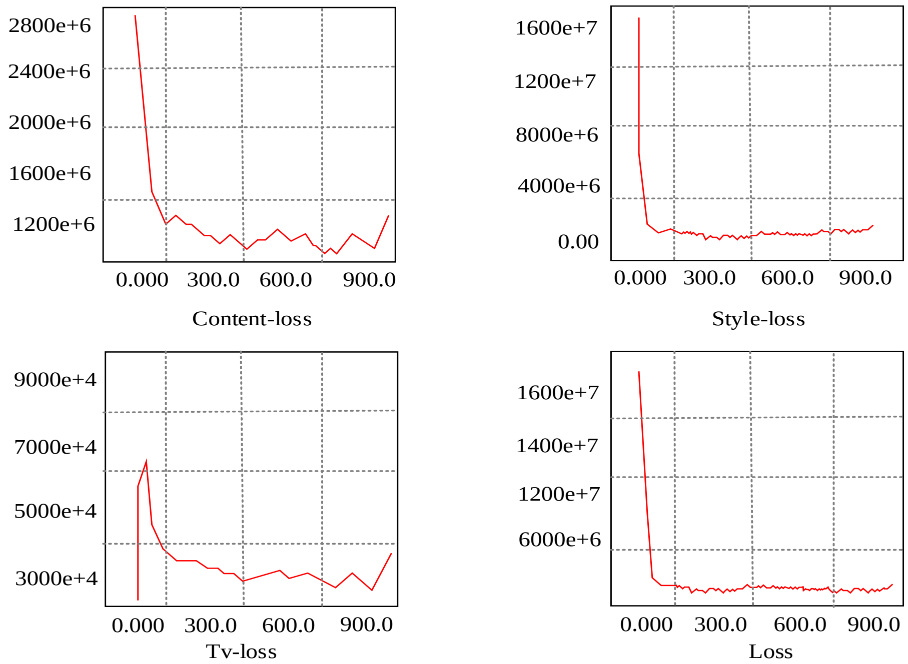


Figure 9: Convergence of image loss in ink painting rendering.

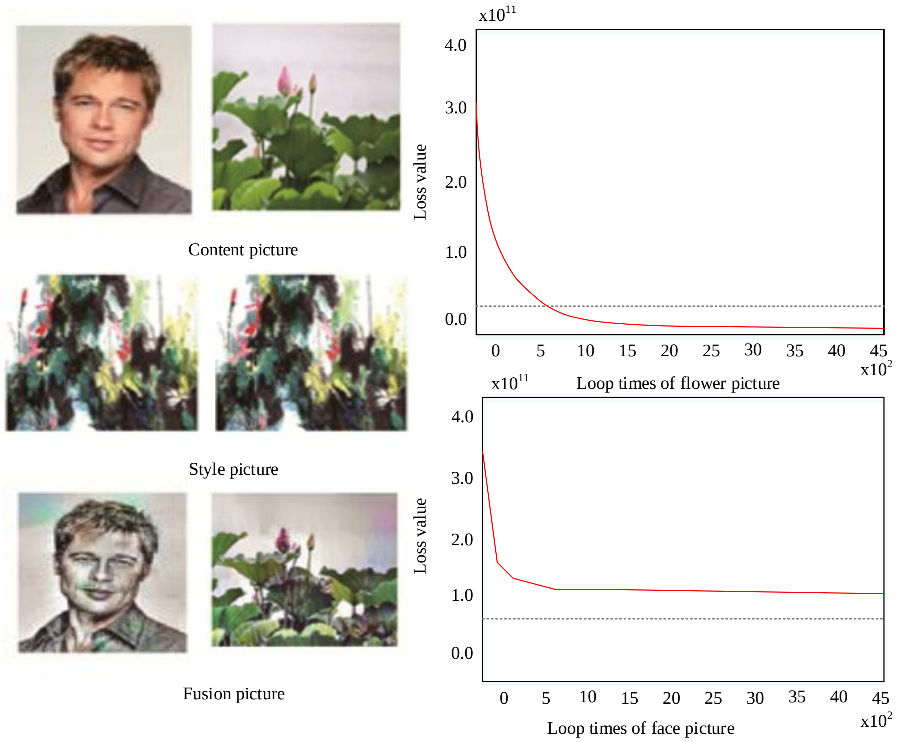


Figure 10: The fusion effect of lotus painting style and real lotus images.

According to Figure 10, when the lotus ink painting style image is fused with a real lotus image, the output image shows a good fusion of content and style, particularly in the lotus image. This demonstrates the achievement of the expected goal of this research. On the other hand, when fusing a real portrait image with the ink painting style image, although an image with ink painting features is obtained, the inherent differences between the portrait and the style image result in a lack of artistic characteristics in terms of fine texture and details. In summary, due to the deeper network layers of the neural network used in this research, it captures more hidden features, resulting in the preservation of more stylistic characteristics and better artistic representation in automatic rendering.

In order to further validate the advantages of the research method, experiments will compare Generative Adversarial Networks (GAN), Style Transfer, and Convolutional Neural Networks (CNN). The performance comparison results of different methods are shown in

Table 5. In terms of accuracy, the research method achieved 95.5%, significantly higher than GAN, Style migration, and CNN. This indicates that the research method has advantages in recognition accuracy. In terms of image quality, the research method is 90.3%, higher than GAN's 85.6%. In terms of satisfaction, the research method is 96.7% ahead of other methods, indicating that the research method can better meet the needs of users. In terms of recognition, the research method demonstrated its superiority at the highest level of 98.3%. Overall, the research method has good performance.

Table 5: Performance comparison results for the different methods.

Method	Accuracy	Picture quality	Degree of satisfaction	Recognition degree
Research method	95.5%	90.3%	96.7%	98.3%
GAN	90.2%	85.6%	92.3%	93.1%
Style migration	86.5%	84.1%	81.3%	90.3%
CNN	88.3%	85.2%	83.4%	91.3%

Table 6: Comparasting results of different methods.

Method	Error rate	Computing time	Memory usage ratio
DTCN	4.6%	560ms	20.3%
BiGAN	5.9%	780ms	26.4%
SRGAN	5.7%	986ms	37.7%

This is due to the advantage of DTCN in capturing sequence features, which can effectively learn the dependency relationships in the input sequence, thereby improving recognition accuracy. DTCN has optimized convolution kernels and pooling strategies, as well as appropriate activation functions. The selection of these parameters helps to reduce noise and distortion while preserving image details, thereby improving image quality. DTCN can adaptively adjust the style of generated images according to user needs, thereby better meeting user aesthetics. In addition, by introducing adaptive adjustment parameters, DTCN can maintain the content and structure of the original image during style transfer, further improving satisfaction. Meanwhile, the adversarial loss function used by DTCN during training can guide the model to generate images that are closer to real data.

To verify the superiority of the proposed DCTN network, the experiment compared Bidirectional GAN (BiGAN) and Super Resolution GAN (SRGAN). The comparison results of different methods are shown in Table 6. The results show that the DCTN network performs well in various indicators. Its error rate is 4.6%, lower than BiGAN's 5.9% and SRGAN's 5.7%; The computation time is 560ms, which has significant advantages compared to BiGAN and SRGAN; The memory utilization rate is 20.3%, significantly lower than BiGAN's 26.4% and SRGAN's 37.7%. This indicates that DCTN networks outperform BiGAN and SRGAN in terms of performance. These advantages enable DCTN networks to have higher efficiency and better performance in practical applications, which have important impacts on fields such as image processing and computer vision, and help improve the accuracy and practicality of related technologies.

6 Discussion

The proposed DTCN in the study has demonstrated significant advantages in multiple aspects and shows significant differences compared to existing technologies. Compared with existing ResNets, DTCN has a lower classification error rate and exhibits better generalization ability. In terms of convergence speed, the training loss function value of DTCN is lower, indicating that its network optimization ability is stronger and its features can be more fully utilized. This indicates that it performs well in feature propagation efficiency and can reach a steady state faster. In terms of artistic output quality, DTCN processed images have lower loss values and higher quality, which can better integrate content and style. For example, the fusion effect of lotus images is good, and more hidden features can be captured in deeper network layers, preserving more style features and artistic expression. In contrast, other methods such as GAN, style transfer, and traditional CNN perform slightly worse on these key metrics. The ability of DTCN to prevent gradient vanishing enables it to better learn useful information during training, which is one of the important reasons for its excellent results and provides better solutions and technical support for the development of related fields such as automatic ink painting rendering. Overall, the proposal of DTCN provides a new solution for automatic ink painting rendering technology. It not only improves the accuracy of classification and accelerates the convergence speed of the network, but also optimizes the quality of artistic output. These advantages make DTCN have broad application prospects in the field of automatic ink painting rendering.

7 Conclusion

With the significant improvement in computational performance, artificial intelligence technologies, represented by CNN, have rapidly developed. In this research, a deep CNN model was constructed to extract the texture features of ink paintings, and these texture features were automatically rendered onto content images, resulting in content images with distinctive ink painting characteristics. Additionally, by comparing the performance of residual networks and triangle networks, it was concluded that the triangle network structure had stronger network optimization capabilities. It effectively and fully utilized features, and converged faster. When the parameter count was 30.1, DTCN demonstrated the shortest testing time and the lowest classification error rate. At this point, the testing time was 1.5 seconds, and the classification error rate was only 3.5%. Furthermore, during the training process, it was observed that the style image and total loss consistently converged rapidly, while the noise image loss experienced a rapid increase, followed by a rapid decrease before gradually converging. The accuracy of the research method reached 95.5%, significantly higher than that of comparative methods such as GAN. The research method is 96.7% ahead of other methods, indicating that the research method can better meet the needs of users. In summary, this research based on deep CNN techniques for automatic ink painting rendering not only obtained output images with distinctive ink painting characteristics but also achieved digital image ink style rendering applications.

The study provides an efficient and stable deep learning method for the transfer of ink painting styles. Its superior performance and high accuracy indicate that this method can better meet the needs of users in artistic creation. In the future, this method is expected to expand to other art fields, including oil painting, watercolor, etc., providing more possibilities for artists to create. Meanwhile, this research achievement also contributes to promoting the application of deep learning technology in cultural heritage protection and traditional art inheritance. Additionally, the potential of DTCN for other style images and the lack of diversity in image styles present opportunities for future research. The main limitation of this study is that the features of ink painting are insufficient in the detailed textures of some output images. How to make the correct features cover more image details is a problem that future research needs to focus on exploring and solving.

References

[1] Wan K, Liu X, Yu J, Zhang X, Du X, Guizani N. (2021). Compiler-based efficient CNN model construction for 5G edge devices. IEEE Transactions on Intelligent Transportation Systems, 22(8):5261-5274. https://doi.org/10.1109/TITS.2021.3056426

[2] Zhang C, Zhu F, Lv Y, Ye P, Wang F. MLRNN: Taxi demand prediction based on multi-level deep learning and regional heterogeneity analysis. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(7):8412-8422. https://doi.org/10.1109/TITS.2021.3080511

[3] Zhang S, Yin R, Zhang M. (2024). Dynamic unstructured pruning neural network image super-resolution reconstruction. Informatica, 48(7): 11-22. https://doi.org/10.31449/inf.v48i7.5332

[4] Wang QA, Li XA, Xu CB, Yan TA, Li YA. (2021). Bubble recognizing and tracking in a plate heat exchanger by using image processing and convolutional neural network. International Journal of Multiphase Flow, 138: 1-13. https://doi.org/10.1016/j.ijmultiphaseflow.2021.103593

[5] Peng Y, Wang X, Zhang J, Liu S. (2021). Pre-training of gated convolution neural network for remote sensing image super-resolution. IET Image Processing, 15(3):1179-1188. https://doi.org/10.1049/ipr2.12096

[6] Thangarajan S K, Chokkalingam A. (2021). Integration of optimized neural network and convolutional neural network for automated brain tumor detection. Sensor Review, 41(1):16-34. https://doi.org/10.1108/SR-02-2020-0039

[7] Deperlioglu O ,Kose U, Gupta D, Khanna A, Giampaolo F, Fortino G. (2022). Explainable framework for Glaucoma diagnosis by image processing and convolutional neural network synergy: Analysis with doctor evaluation. Future Generation Computer Systems. 129: 152-169. https://doi.org/10.1016/j.future.2021.11.018

[8] Ghosh S K, Ghosh A. (2021). A novel retinal image segmentation using rSVM boosted convolutional neural network for exudates detection. Biomedical Signal Processing and Control, 68(2):1-13. https://doi.org/10.1016/j.bspc.2021.102785

[9] Deshpande A M, Patale S R, Roy S. (2021). Removal of line striping and shot noise from remote sensing imagery using a deep neural network with post-processing for improved restoration quality. International journal of remote sensing, 42:7357-7380. https://doi.org/10.1080/01431161.2021.1957512

[10] Xie X, Lv B. (2021). Design of painting art style rendering system based on convolutional neural network. Scientific Programming, 2021, (8): 1-11. https://doi.org/10.1155/2021/4708758

[11] Platkevi A, Curtis C, Skora D. Fluidymation: Stylizing animations using natural dynamics of artistic media. Computer Graphics Forum, 40(7): 21-32. https://doi.org/10.1111/cgf.14398

[12] Yao W, Sohail M. (2022). Design of artistic creation style extraction model based on color feature data. Mathematical Problems in Engineering, 1: 1-9. https://doi.org/10.1155/2022/4811191

[13] Zhang, X, Demiris, Y. (2023). Visible and infrared image fusion using deep learning. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(8), 10535-10554. https://doi.org/10.1109/TPAMI.2023.3261282

[14] Zheng P, He H, Gao Y, Tang P W, Wang H L,

Peng J, Wang L, Su C M, Ding S Y. (2022). Speeding up the topography imaging of atomic force microscopy by convolutional neural network. Analytical chemistry, 94(12):5041-5047. https://doi.org/10.1021/acs.analchem.1c05056

[15] Wy A, Gl A, Zb Ling, Wy A. (2020). Convolutional neural network with feature reconstruction for monitoring mismatched photovoltaic systems. Solar Energy, 212(3):169-177. https://doi.org/10.1016/j.solener.2020.09.026

[16] Deeba F, Zhou Y, Dharejo F A, Khan M A, Das B, Wang X Z, Du Y. (2021). A plexus-convolutional neural network framework for fast remote sensing image super-resolution in wavelet domain. IET image processing, 15(8):1679-1687. https://doi.org/10.1049/ipr2.12136

[17] Philip F M, Jayakrishnan V, Ajesh F, Ajesh F, Haseena P. (2021). Video anomaly detection using the optimization-enabled deep convolutional neural network. The Computer Journal, 2021, 65(5):1272-1292. https://doi.org/10.1093/comjnl/bxaa177

[18] Liu Y, Pu H, 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, 2021, 113(7):193-204. https://doi.org/10.1016/j.tifs.2021.04.042

[19] Kadhim A R, Khudeyer R S, Alabbas M. (2024). Facial sentiment analysis using convolutional neural network and fuzzy systems. Informatica, 48(12): 15-32. https://doi.org/10.31449/inf.v48i12.6151

[20] Li H, Zhu W. (2024). Art image style conversion based on multi-scale feature fusion network. Informatica, 48(10): 1-18. https://doi.org/10.31449/inf.v48i10.5960

[21] Hou Y, Fan H, Li L, Li B. (2021). Adaptive learning cost: ensitive convolutional neural network. IET Computer Vision, 15(5):346-355. https://doi.org/10.1049/cvi2.12027