

Art Style Classification Based on Spatial Downscaling and Convolutional Neural Network

Jie-Lan Lei*

Hefei Preschool Education College, Hefei 230013, P. R. China
19855151972@163.com

Bei-Jing Tao

Hefei Preschool Education College, Hefei 230013, P. R. China
13637065831@163.com

Baral Abinash

Joyride Technology Co., Ltd., Kingston K7P1R1, Ontario, Canada
mechbaral@gmail.com

*Corresponding author: Jie-Lan Lei

Received May 11, 2024; revised September 16, 2024; accepted January 3, 2025.

ABSTRACT. *Intending to the issue that current art style classification algorithms ignore important and redundant features of images, which leads to low classification efficiency, this article designs an art style classification method based on spatial downscaling and Convolutional Neural Network (CNN). Firstly, for the issue that the weights of CNN cannot be updated, a new activation function IReLU is suggested on the basis of ReLU and SoftPlus to enhance the anti-interference ability of the model. On this basis, the CNN is adopted to comprehensively extract the spatio-temporal features of the oil painting image style, and the extraction of important features is enhanced through the interaction of the attention mechanism and the residual network, and then an optimal projection space is obtained by using the Linear Discriminant Analysis (LDA) spatial dimensionality reduction algorithm, and the spatio-temporal features of the oil painting image are downscaled by the projection matrix in the obtained subspace, and the mapped low-dimensional features serve as the input to the softmax classifier input, which in turn improves the efficiency of classification. Finally, it is tested on the Pandora18k dataset and compared with the other three methods, the accuracy and F1 of the suggested method are improved by at least 10.8% and 6.4%, respectively. The experimental outcome implies that the algorithm effectively improves the performance of important features for the task of oil painting art style classification and has good practical value.*

Keywords: Spatial dimensionality reduction; Convolutional neural network; Activation function; Attention mechanism; Residual network

1. **Introduction.** The style of an artwork is an important part of its meaning, and the understanding and categorization of style is an important issue in the quantification of art [1]. Oil painting is the main painting method in the history of Western painting, with the development of digital technology, many museums are constructing digital oil painting libraries, which achieves a new era of art sharing given by digitalization. For oil paintings, the artistic style is not an abstract and empty existence, but is realized in concrete works. When painters paint, they harmonize the content and form of the picture to show the ideological tendency and artistic characteristics of the school or painter, thus forming their unique artistic style [2, 3, 4]. Therefore, how to correctly identify the artistic style

of oil paintings is the key to establish a large-scale art database. In the field of computers, many researchers have invested considerable efforts to classify the styles of oil paintings, and to realize the quantification of art by using scientific classification means as a tool [5, 6]. By classifying digital oil paintings according to their themes, styles, colors and other characteristics, social networking and search platforms can use the automatic classification technology of digital oil paintings to help users find their interesting content more easily.

1.1. Related work. In the traditional task of oil painting style classification, researchers usually use computer vision and machine learning technology to extract and analyze the characteristics of oil painting such as color, texture and brush strokes. These characteristics can help to identify different styles of oil paintings, such as abstraction, realism, impressionism and so on. By modeling and classifying these features, researchers can realize automatic identification and classification of oil painting styles. Abry et al. [7] used various supervised and unsupervised learning methods to classify malefactors and developed a system to extract Van Gogh's distinctive brushstrokes. Liong et al. [8] proposed a metric learning approach to find out the paths of influence between different paintings. Georgoulaki [9] studied the classification of paintings based on texture features and extracted the features of LBP, LPQ and LCP and fed them into an SVM model for classification. Jiang et al. [10] proposed the edge size histogram feature, which is combined with the traditional underlying features such as the color histogram, to identify the Chinese paintings from all the painting images. Gao et al. [11] proposed an algorithm to classify Chinese paintings based on the content of the picture, using the grayscale covariance matrix to extract local patterns and alignment rules. Amelio and Pizzuti [12] proposed a weighted nearest-neighbor inheritance algorithm to classify oil paintings using multidimensional features. Deng et al. [13] used color statistics as global features of paintings, segmentation of objects in paintings as local features, and unsupervised learning method to cluster the features to classify paintings.

In the deep learning painting style classification, Kim et al. [14] used an artificial neural network approach to achieve automatic classification of oil painting styles, and they achieved 54.5% accuracy in WikiPainting over more than 25 painting styles. Ao et al. [15] proposed a feature extraction method using K-means clustering algorithm to cluster based on these features to classify art styles. Yang and Min [16] proposed a deep convolutional network-based oil painting classification method to visualize low-level to high-level features. Chen and Yang [17] used BP neural network to classify oil painting styles based on a deep learning network model for oil painting classification. Wu et al. [18] proposed to use a deep residual neural network to solve the problem of detecting the artistic style of a painting. Albadarneh and Ahmad [19] classified images based on features obtained from a KNN neural network pre-trained for an image classification task and predicted the painting style. Jiang and Yang [20] used a Convolutional Neural Network (CNN) to separate the painting style from the content of the painting and recoded it so as to achieve the classification of oil paintings without using different styles. However, in the process of oil painting image processing, feature extraction produces a large amount of redundancy and consumes a large amount of computational resources. In order to improve the computational performance, scholars began to study how to reduce the redundant features of the image. Chen et al. [21] proposed an algorithm to describe the features of oil paintings by using PCA, and applied the PCA dimensionality reduction method to the feature extraction of oil paintings, which effectively reduces the number of feature dimensions of the image. Aubry et al. [22] proposed a method for classifying the images of paintings based on the Linear Discriminant Analysis (LDA) method, which takes the image neighboring pixel with similar features between them are considered in the classification model.

1.2. Contribution. In existing studies, the oil painting image style classification method based on CNN significantly improves the classification performance, however, there is a bottleneck in extracting important features, which leads to low classification efficiency. To address this issue, this article designs an art style classification method based on spatial dimensionality reduction and CNN. Firstly, to deal with the issue that the negative region of the ReLU activation function in CNN takes the value of 0 and the corresponding weights cannot be updated, a new activation function IReLU is proposed on the basis of ReLU and SoftPlus, which solves the issue that it is hard to manually adjust the parameters of ELU and Leaky-ReLU. On this basis, the CNN is utilized to extract spatio-temporal features by fusing temporal and spatial information, and through the interaction of the attention mechanism and the residual network, more attention is paid to the important channels and specific regions, which improves the discriminative ability of the features and enhances the ability of perceiving the spatio-temporal information. Then the features are mapped into a low-dimensional subspace by LDA algorithm, and the mapped low-dimensional features are used as inputs to softmax classifier, which in turn improves the efficiency of classification.

2. Theoretical analysis.

2.1. Spatial downscaling technique. The spatial downscaling technique is to learn the training samples by projecting the original features to another space, where the new features remove the redundant information and retain the key information. The LDA algorithm [23] takes the extreme value of the objective function and projects the original features to a desired low-dimensional space by the obtained projection matrix, thus transforming the classification issue to that low-dimensional space.

Assume that there are m training samples $x_1, x_2, \dots, x_m \in R^c$, and that the training set has a total of M classes: d_1, d_2, \dots, d_M , and assume that there are m_i training samples in class i .

Define n to be the mean vector of all samples, n_i to be the mean of the samples of category i , and the scatter matrix between samples in the same category to be S_v and between samples of different categories to be S_a .

$$S_v = \sum_{i=1}^M \sum_{x \in D_i} (x - n_i)(x - n_i)^T \quad (1)$$

$$S_a = \sum_{i=1}^M m_i(n_i - n)(n_i - n)^T \quad (2)$$

The sample x is projected by the optimal projection matrix v to obtain the sample $y = v^T x$. The scatter matrix between samples of the same category after projection is S_v^* ($S_v^* = v^T S_v v$), and the scatter matrix between samples of different categories is S_a^* ($S_a^* = v^T S_a v$).

After projection through the projection matrix, the samples of the same class are concentrated as much as possible and the samples of different classes are separated as much as possible. The objective function is defined as follows.

$$J(v) = \frac{v^T S_a^* v}{v^T S_v^* v} \quad (3)$$

2.2. Convolutional neural network. CNN have been favored by many researchers in the field of image processing due to their ability to automatically capture deep features in the input data without any human supervision [24]. This article focuses on improving the classification performance based on CNN. A traditional CNN is a trainable multilayer framework consisting of multiple feature extraction stages, including convolutional, non-linear and pooling levels, and finally a fully connected layer and a classifier. The diagram of CNN is implied in Figure 1.

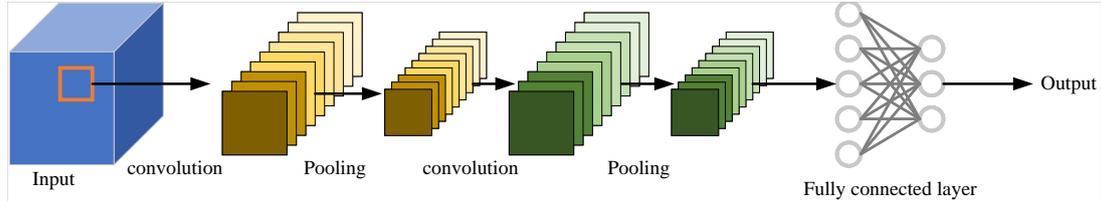


Figure 1. Schematic diagram of CNN

(1) Convolution level. When performing convolution, the convolution kernel moves according to a predefined Stride, performs matrix dot products on the scanned area, superimposes the bias values, and weights and sums them to obtain the output value at the corresponding position. The general convolution operation is implied below.

$$Y = conv \left(\sum_{m=0}^{F_W-1} \sum_{n=0}^{F_H-1} V_{m,n} x_{i+m,j+n} \right) + A = conv(V, X) + A \quad (4)$$

where X denotes the input matrix, Y denotes the output matrix, F_W and F_H denote the width and height of the convolution kernel respectively, V denotes the weight matrix.

(2) Pooling level. Within a region $R_{m,n}^c$, the maximum activity value is taken for all neuron activity values, which corresponds to the following formulas, respectively.

$$Y_{m,n}^c = \max_{i \in R_{m,n}^c} (x_i) \quad (5)$$

where x_i is the activation value of each neuron in region $R_{m,n}^c$.

(3) Fully connected layer. It mainly realizes the goal of classification. In the case of levels $n - 1$ and n , any neuron node in level $n - 1$ is connected to all the nodes in level n . At the same time, the computation integrates each node in level n . When the activation function is computed, the input features are weighted by all the nodes in level $n - 1$. The activation function of level n is the weighting of all the nodes in level $n - 1$.

3. Optimization of activation functions for CNN. Activation function is an indispensable part of CNN model, which is an important factor to improve the performance of the model and enhance the ability of the model to deal with complex problems. From the early saturated activation function Sigmoid, Tanh to the latest non-saturated activation function Mish, Swish, etc., the activation function has made great progress. Focusing on the issues of gradient disappearance, neuron death and mean shift, which are common in the current activation functions, this paper constructs IReLU on the basis of ReLU and SoftPlus, which is a high-efficiency activation function that improves the performance of the model but does not introduce additional hyperparameters.

To solve the issue of gradient dispersion in the nonlinear saturated activation functions Sigmoid and Tanh of CNN, Chen and Ho [25] proposed the ReLU function by mimicking the mechanism of neuron excitation in the human brain, and its mathematical form is as follows.

$$f(x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (6)$$

SoftPlus is a smoothed approximation of ReLU, which allows the model to maintain sparsity while ensuring that a small amount of negative regional information is passed through the model. The mathematical expression is as follow:

$$f(x) = \ln(1 + e^x) \quad (7)$$

The following IReLU activation functions are constructed based on the two types of activation functions in Equation (6) and Equation (7).

$$f(x) = \begin{cases} x, & x > 0 \\ x(e^x + x \ln(1 + e^x)), & x \leq 0 \end{cases} \quad (8)$$

From Equation (8), it can be seen that when $x > 0$, the IReLU function does a constant mapping to the input features, i.e., the input value is equal to the output value; and when $x \leq 0$, the output result is a linear combination of ReLU and SoftPlus. Compared with ReLU, IReLU adds negative sample information to the model during the transfer process. Compared with SoftPlus, which is always greater than 0, IReLU has both positive and negative values, which can effectively mitigate the "mean shift" problem.

Assuming that the input value of each neuron is z and the output result after the activation function is b , and the activation function is δ , the following relation is given.

$$z = vx + a, \quad b = \delta(z) \quad (9)$$

Now take the network structure of levels $i - 1$, i and $i + 1$ in the CNN, and let the weight of the l -th neuron in level $i - 1$ to the j -th neuron in level i of the neural network be v_{jl}^i , the bias from the neuron in level i to the j -th neuron be a_j^i , and the input to the j -th neuron in level i be z_j^i , the output result is b_j^i .

$$b_j^i = s(z_j^i) = s\left(\sum_{l=1}^m b_l^{i-1} v_{jl}^i + a_j^i\right) \quad (10)$$

where m is the number of neurons in layer $i - 1$.

When f is an IReLU function, there is:

$$b_j^i = \begin{cases} z_j^i, & z_j^i > 0 \\ z_j^i \left(e^{z_j^i} + z_j^i \ln(1 + e^{z_j^i}) \right), & z_j^i \leq 0 \end{cases} \quad (11)$$

From the above equation, the difference between IReLU and ReLU is that in the forward propagation of the model, when the input value is less than or equal to 0, the output result is no longer a fixed constant zero, which makes the network produce a small gradient flow in the back propagation process, which promotes the updating of the weights and biases in the network, and thus reduces the error rate of the model.

4. Art style classification based on spatial downscaling and CNN.

4.1. CNN-based stylistic feature extraction for oil painting images. Intending to the existing oil painting image style classification methods which are difficult to effectively extract the key features of the style and lead to inefficient classification, this article designs an art style classification method based on spatial downscaling and CNN, which is shown as Figure 2. Firstly, CNN is used to comprehensively extract the spatio-temporal features of oil painting image style, and through the interaction of attention mechanism and residual network, it pays more attention to the important channels and specific regions to improve the discriminative ability of the features, and then the features are mapped into a low-dimensional subspace by LDA algorithm, and the mapped low-dimensional features are used as inputs of softmax classifier, which in turn improves the efficiency of classification.

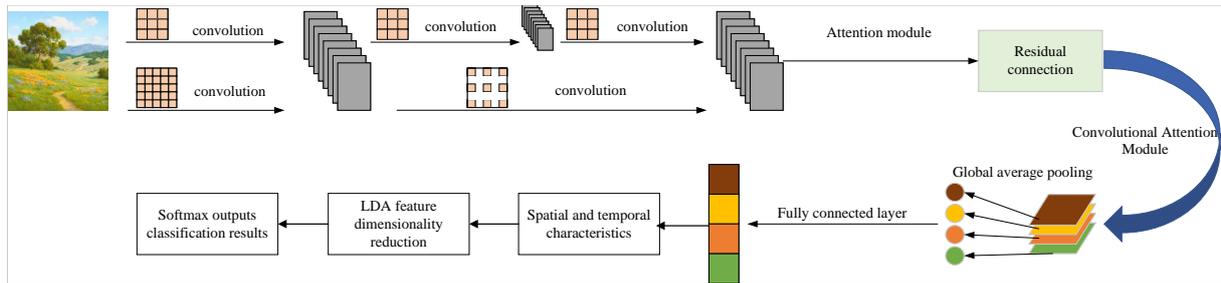


Figure 2. The model of the suggested method for categorizing oil painting styles

Assuming that the input to the network is any oil painting image $X \in R^{M \times T \times D \times G \times V}$, M and T denote the number of training samples and the input oil painting pixels, respectively. D , G and V denote the number of channels, the length and the width of the feature map, respectively. In order to reduce the computational effort of the network, the input channels of the painting image are downscaled using 2D convolution.

$$X^S = Conv_1^* X, \quad X^S \in R^{M \times T \times D/R \times G \times V} \quad (12)$$

where $Conv_1^*$ is the convolution operation; S is the multiplicity of dimensionality reduction; and X^S is the image after dimensionality reduction.

The temporal splitting operation is performed on the dimensionality reduced image.

$$Split(X^S) = [X^S(1), X^S(2), \dots, X^S(T)] \quad (13)$$

Considering the problem of inconsistent local stylistic features in oil painting images, the fusion of channel dimension and temporal sequence is used to extract temporal features.

$$MX(t) = Diff(Conv_{DW} * X^S(t+1) - X^S(t)) \quad (14)$$

where $MX(t)$ is the temporal feature at time t , $Diff$ is the difference operation; $Conv_{DW}^*$ is the channel-by-channel convolution operation.

The oil painting feature map with local spatial information is then channel interacted using a 2D convolution and the dimension is transformed to $[M \times T \times D/R \times GV]$. The global spatial feature with channel dimension 1 is also obtained using a 2D convolution and the dimension is transformed to $[N \times T \times GV \times 1]$. Then, the feature map with local spatial information and the global spatial feature map are dot-multiplied to obtain the channel descriptor that combines the local and global spatial information. Then, the feature map with local spatial information and the global spatial feature map are dot-multiplied to obtain a channel descriptor that combines the local and global spatial information.

$$M_{MX}^R = Trans(Conv_1 * M_{MX}^R) \odot IReLU(Conv_1 * M_{MX}^R) \quad (15)$$

where $M_{MX}^R \in R^{M \times T \times D/R \times 1 \times 1}$ denotes the channel descriptor; $Trans(\cdot)$ denotes the dimension transformation operation, \odot denotes the dot product operation, and IReLU is the optimized activation function.

A two-dimensional convolution operation is performed on M_{MX}^R to restore its dimension to the original number of channels D and to obtain the normalized modulation weights for each channel.

$$A_{Att,D} = IReLU(Conv_1 * M_{MX}^R) \quad (16)$$

The final feature map is obtained by multiplying the channel modulation weights $A_{Att,D}$ with the original input feature map points, which accomplishes the excitation of heavy channels in the feature map and enhances the differentiated features. Considering that some key features may be gradually suppressed, the residual connection [26] is adopted to stimulate the important channels while preserving the static information.

$$X^* = X + A_{Att,D} \odot X \quad (17)$$

Based on the above residual connections, the oil painting features are adaptively enhanced by channel attention module $A_C(\cdot)$ and spatial attention module $A_S(\cdot)$. $A_C(\cdot)$ aims to assign adaptive weights to different feature channels. Given an input feature map $X \in R^{M \times T \times D \times G \times V}$, it is first compressed by average pooling and maximum pooling to obtain average pooled feature $AP(X)$ and maximum pooled feature $MP(X)$ respectively.

Then, these two features are forwarded to a shared network $F(\cdot)$ consisting of two fully connected layers to capture the internal relationship between the feature channels. Finally, the channel attention vector $A_C(X) \in (0, 1)^{1 \times c}$ is obtained by the IReLU activation function, which is expressed as follows.

$$A_C(X) = IReLU(F(AP(X)) + F(MP(X))) \quad (18)$$

The spatial attention module $A_S(\cdot)$ is able to assign different weights to feature regions. In this module, the features generated by average pooling and maximum pooling are connected to a common convolutional layer to explore spatial relationships. The spatial attention graph $A_S(X) \in (0, 1)^{G \times V \times 1}$ can be represented as follows.

$$A_S(X) = IReLU(Conv([AP(X); MP(X)])) \quad (19)$$

Finally combining $A_C(X)$ and $A_S(X)$ yields the enhanced oil painting style characterization shown below.

$$X' = X^* + (A_C(X) + A_S(X)) \odot MX(t) \quad (20)$$

4.2. Feature downscaling and stylistic classification of oil painting images. After obtaining the stylistic features of each oil painting image, this paper maps its features into a low-dimensional subspace through $X_i = W^T Y_i$. $W \in R^{(n \times m) \times a}$ is the solution matrix, where a is the dimension of the space. The low-dimensional features mapped by the LDA algorithm are used as inputs for the subsequent classification of oil painting images, which improves the efficiency of the classification.

Suppose there are n samples of oil painting images, which have a total of N categories: d_1, d_2, \dots, d_N . The goal of classification of oil painting images is to find a projection matrix W such that the projection of oil painting images on this projection matrix minimizes the local scatter matrix between samples of the same category and maximizes the local scatter

matrix between samples of different categories. According to the definition in Section 2.1, the local scatter matrices between oil painting images of the same category and different categories are found to be A and C , respectively.

$$A = \sum_{q=1}^N \sum_{Y_i \in d_q} \sum_{Y_j \in d_q} \frac{1}{n_q} \frac{1}{N_i} S_{i,j} (Y_i - Y_j)(Y_i - Y_j)^T \quad (21)$$

$$C = \sum_{q=1}^N \sum_{Y_i \in d_q} \sum_{Y_j \notin d_q} \frac{1}{n_p} \frac{1}{M_i} (Y_i - Y_j)(Y_i - Y_j)^T \quad (22)$$

where N_i denotes the number of similar points of oil painting Y_i . M_i denotes the number of dissimilar points of oil painting Y_i . $S_{i,j}$ denotes the similarity between oil painting Y_i and oil painting Y_j as follows.

$$S_{i,j} = 1 - S_{i,j}^q S_{i,j}^t \quad (23)$$

where $S_{i,j}^t$ and $S_{i,j}^q$ denote the temporal and spatial characteristics of the oil painting style, respectively, as indicated below.

$$S_{i,j}^p = 1 - \exp\left(-\frac{\|Y_i' - Y_j'\|_2^2}{t_1}\right) \quad (24)$$

$$S_{i,j}^t = 1 - \exp\left(-\frac{\|MY_i^t - MY_j^t\|_2^2}{t_2}\right) \quad (25)$$

where t_1, t_2 correspond to the parameters of the Gaussian function for the temporal and spatial features, respectively.

$$\hat{A} = A + A^* \quad (26)$$

where A is defined as in Equation (21) and A^* is defined as follows:

$$A^* = \sum_{q=1}^N \sum_{y_i \in d_q} \sum_{y_j \in N(Y_i)} \frac{1}{n_q} \frac{1}{K} S_{i,j} (Y_i - Y_j)(Y_i - Y_j)^T \quad (27)$$

where $N(Y_i)$ is the set of K test neighbors of Y_i .

Thus, the objective function for oil painting image style classification is defined as follows:

$$W = \arg \min_W \text{tr} \left(W^T (\hat{A} + C)^{-1} C W \right) \quad (28)$$

For the issue of solving Equation (28), it can be transformed into Equation (29).

$$(\hat{A} + C)W_j = \beta_j C W_j, \quad j = 1, 2, \dots, p \quad (29)$$

To find the projection matrix W , the matrix $(\hat{A} + C)^{-1} C$ is first eigenvalued decomposed, and the eigenvectors corresponding to the p largest eigenvalues are taken and combined to form the projection matrix W . Eventually, the dimensionally reduced feature X_i can be obtained.

In the predictive classification stage, X_i is used as an input to the softmax's classifier to classify the oil painting image styles, and the value with the highest score probability is the category label value as follows:

$$\hat{z}_i = \max_{d_j \in D} \{X_i \otimes d_j\} \quad (30)$$

where \hat{z}_i denotes the predicted value. In order to make the predicted labels more similar to the true labels, the classifier is trained using the cross-entropy loss [27] of all known class nodes as follows:

$$l = \sum_{i=1}^s z_i \log f(\hat{z}_i) + (1 - z_i) \log f(1 - \hat{z}_i) \quad (31)$$

where z_i represents the true label, $f(\cdot)$ represents the IReLU activation function.

5. Performance testing and analysis.

5.1. Classification performance analysis. The hardware and software environments used for the experiments in this paper are: Ubuntu 24.04.1 LTS Linux operating system, Intel (R) Core (TM) i7-9750H CPU @ 2.60GHz processor, and 16GB of RAM. The experimental program was implemented using the Python programming language. Cross-entropy is used as the loss function in the training process, and the Adam optimizer is used to update the iteration, with the initial learning rate adjusted to 0.001 and the batch size set to 32.

To validate the effectiveness of the algorithm, we choose the public oil painting dataset Pandora18k [28] as the dataset of this paper. This dataset is mainly divided into 15 style labels, such as Impressionism, Abstraction, Surrealism and so on. In order to eliminate the randomness of the experiments and ensure the validity of the experimental results, the data set is run 100 times, and each experiment only determines the number of training and testing sets, randomly selects the images, and takes the optimal result among the 100 times as the final result.

To evaluate the performance of the oil painting classification model more objectively, this paper chooses accuracy, precision, recall, F1, AUC (Area Under ROC) [29] as the evaluation indexes. In order to facilitate the comparative analysis, literature [17] is denoted as ANM_BP, literature [19] as MLO_KNN, literature [20] as EMO_CNN, and this article's method is denoted as ASD_CNN. The comparison results of the accuracy rates are implied in Figure 3, whereas the four models have basically the same initial state, the accuracy rate of ASD_CNN reaches more than 80% in the first place, and when the number of traversals is greater than 50, the model tends to converge, while ANM_BP and MLO_KNN show the same trend, i.e., the accuracy rate increases with the change of learning rate and fluctuates a lot, and EMO_CNN model performs better, and the accuracy rate is only second to that of ASD_CNN, which reaches 80%. Therefore, ASD_CNN has the best classification effect, which not only has high accuracy and fast convergence speed, but also has small fluctuation of the model.

The classification results of the four models are implied in Table 1. ASD_CNN outperforms the comparison models in all classification performance indexes, and improves 14.93%, 14.19% and 6.4% in F1 compared to ANM_BP, MLO_KNN and EMO_CNN respectively. This is due to the fact that ANM_BP and MLO_KNN extract only some of the style features of the oil painting and do not remove the redundant features in the process of classification, which leads to the inefficiency of the classification. The EMO_CNN method only utilizes the CNN to extract the global style features of the oil painting, and does not enhance the extraction of the key features. In contrast, ASD_CNN not only enhances the extraction of important regional features through the attention mechanism, but also

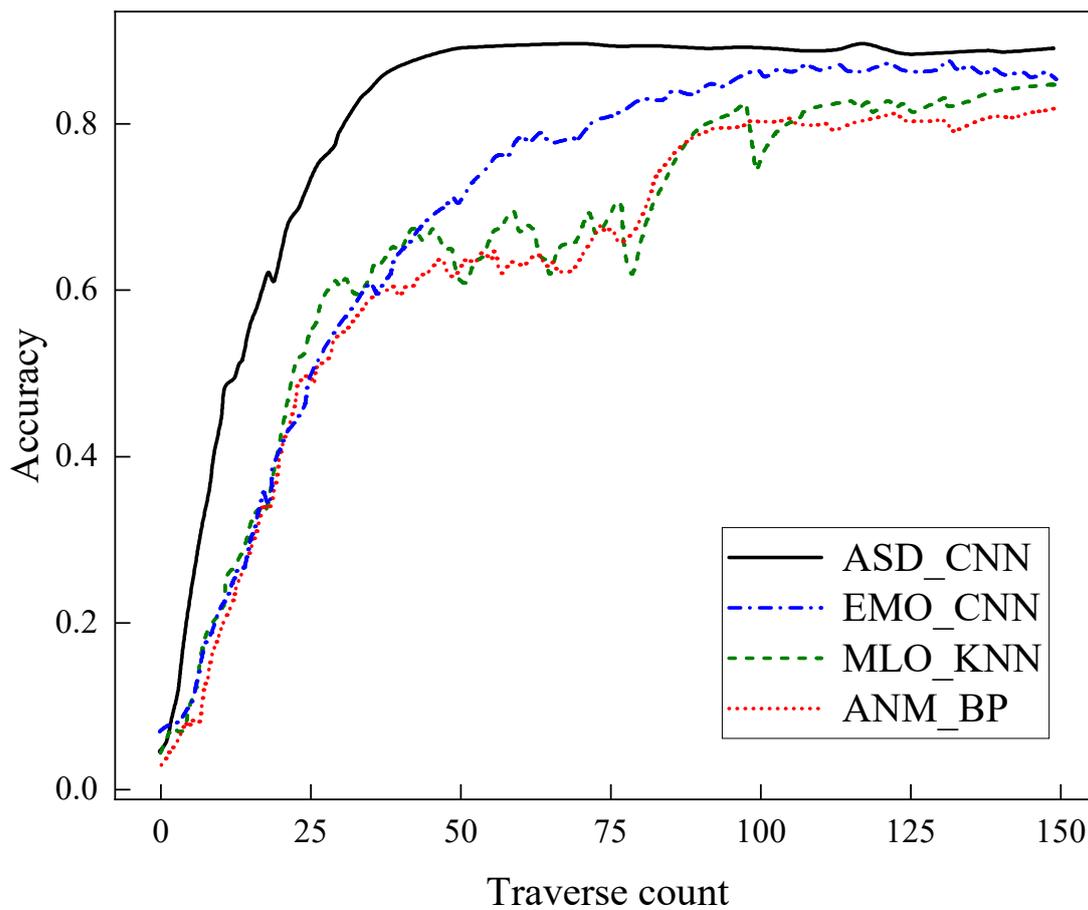


Figure 3. Comparison of classification accuracy

reduces the dimensionality of the features through the LDA algorithm, which in turn improves the classification efficiency. AUC denotes the area from (0, 0) to (1, 1) under the ROC curve, and the closer the AUC is to 1, the better the model performs. ASD_CNN's AUC is the closest to 1, followed by EMO_CNN's performance, and the performance of ANM_BP and MLO_KNN deviate the farthest from 1. Therefore, the ASD_CNN performs the best in classification by combining all the indicators.

Table 1. Comparison of experimental results of classifying oil painting styles

Method	Precision/%	Recall/%	F1 /%	AUC
ANM_BP	73.16	74.35	73.75	0.7129
MLO_KNN	76.24	72.81	74.49	0.7357
EMO_CNN	82.94	81.63	82.28	0.7912
ASD_CNN	89.87	87.52	88.68	0.9175

5.2. Classified losses and time comparison. To further illustrate the classification effectiveness of the ASD_CNN method, the change of loss values with the number of iterations during model training comparing ASD_CNN, ANM_BP, MLO_KNN and EMO_CNN is shown in Figure 4. It can be seen that the loss values of ASD_CNN method are lower than those of ANM_BP, MLO_KNN and EMO_CNN, and converge when the Epochs reaches 60. The loss value of ASD_CNN method decreases to about 1.6 after 100 iterations, while that of EMO_CNN method decreases to about 2 only, and fluctuates greatly

in the later stage, and the loss value of ANM_BP decreases to 3, while that of EMO_CNN method decreases to 3. The loss value of ANM_CNN method decreases to about 2 with large fluctuation in the late stage, and the loss value of ANM_BP decreases to about 3 with large fluctuation, and the loss value of MLO_KNN method decreases to about 4.5 with large fluctuation in the early stage.

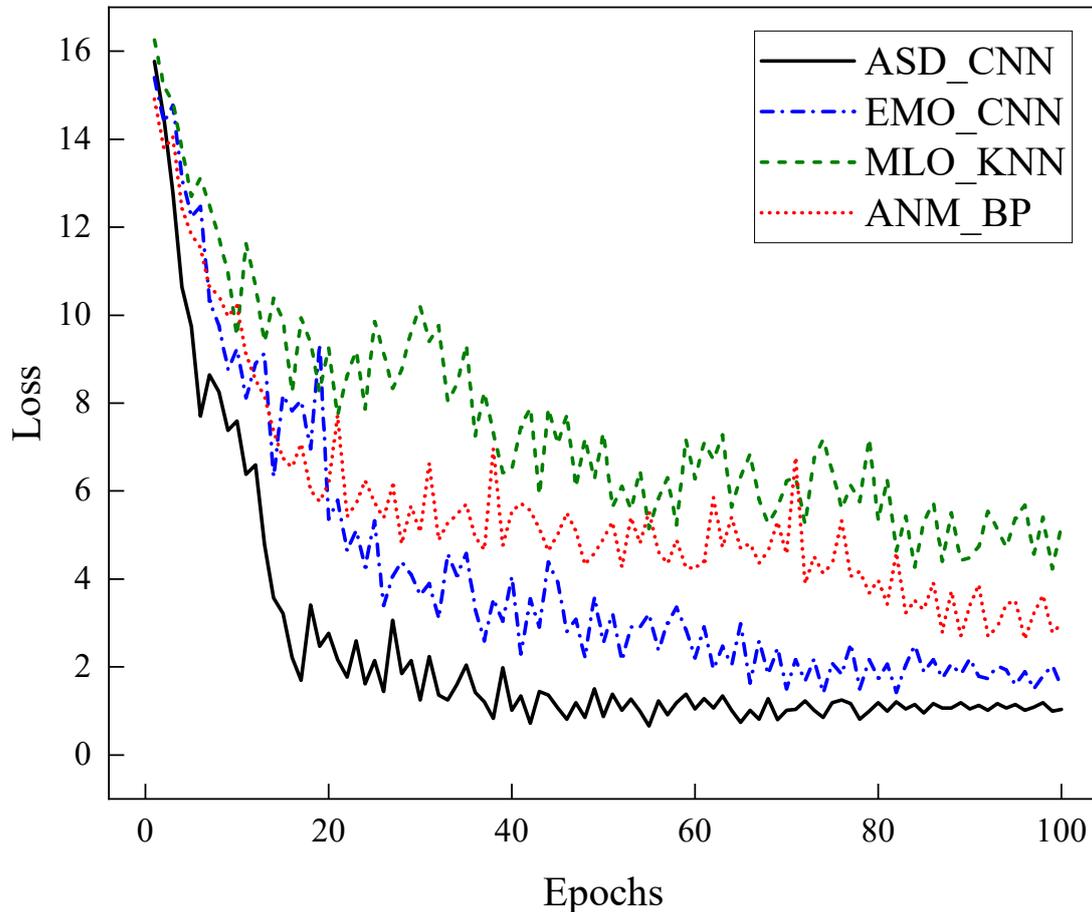


Figure 4. Comparison of classification time of various oil painting style classification methods

The comparison of the classification times of different methods is shown in Figure 5. The shorter the classification time is, the better the classification effect is, and the longer the classification time is, the worse the classification effect is. When the number of oil painting images is 60, the classification times of ASD_CNN, ANM_BP, MLO_KNN and EMO_CNN are 4.9 s, 8 s, 7.8 s and 5.5 s. Therefore, the classification time of ASD_CNN is the lowest among the three methods. This is mainly due to the fact that ANM_BP and MLO_KNN do not consider the key features and do not define the artistic style of oil paintings in terms of their overall senses, and thus the classification effect is poor. The EMO_CNN only utilizes the CNN neural network to study the artistic style of oil paintings, but does not consider the redundancy of the features, resulting in a long classification time. ASD_CNN extracts spatio-temporal features of oil painting styles based on CNN and enhances key features by using the attention mechanism to differentiate different factions of oil painting styles, which has never been considered in the study of oil painting styles, and thus achieves the best classification results.

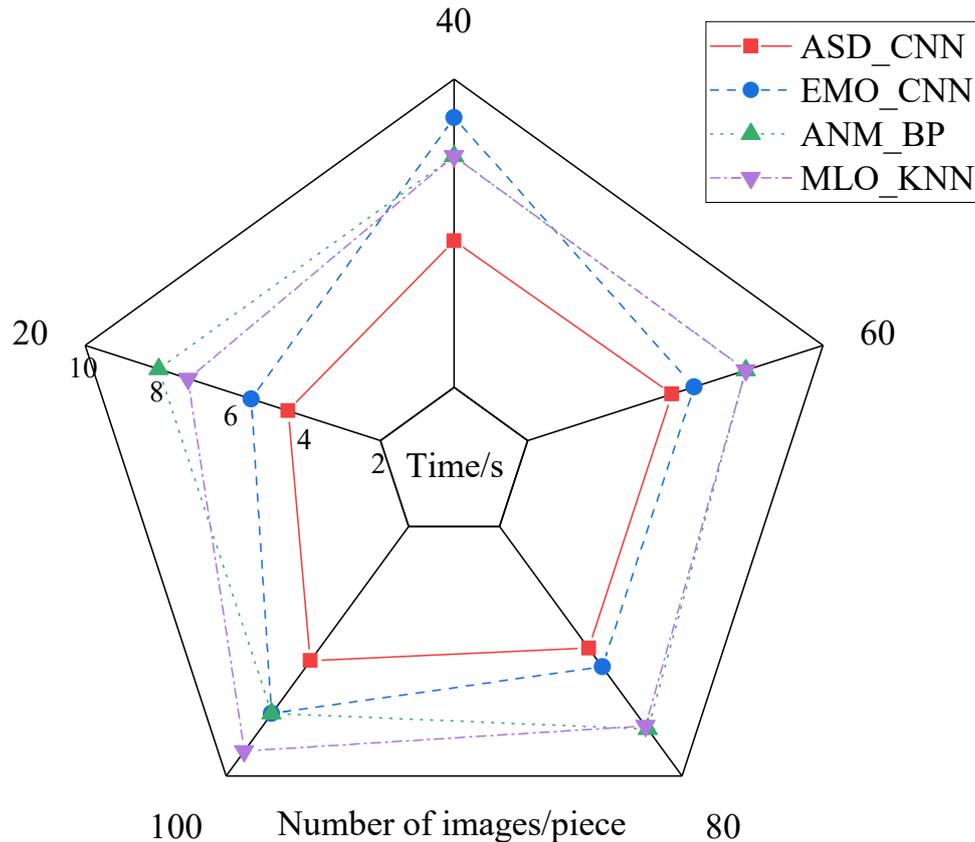


Figure 5. Comparison of oil painting styles categorized over time

6. Conclusion. Intending to the existing oil painting image style classification method is difficult to effectively remove feature redundancy, resulting in the issue of inefficient classification. This article designs an art style classification method based on spatial dimensionality reduction and CNN. First of all, the ReLU activation function in CNN is optimized, and a new activation function IReLU is proposed on the basis of ReLU and SoftPlus. The values of IReLU in the negative region have both positive and negative values, which can effectively alleviate the problem of "mean shift" and improve the anti-jamming ability of the model. On this basis, the spatial and temporal features are extracted by fusing temporal and spatial information with CNN, and through the interaction of attention mechanism and residual network, more attention is paid to the important channels and specific regions, so as to improve the discriminative ability of the features and to enhance the ability to perceive the spatial and temporal information. Then the features are mapped into a low-dimensional subspace by the LDA algorithm, and the mapped low-dimensional features are used as the input of the softmax classifier, which in turn improves the efficiency of classification. Finally, the experimental outcome indicates that the suggested method can effectively improve the accuracy, precision, recall, F1 and AUC values of oil painting style classification, and has better classification performance.

REFERENCES

- [1] N. Van Noord, E. Hendriks, and E. Postma, "Toward discovery of the artist's style: Learning to recognize artists by their artworks," *IEEE Signal Processing Magazine*, vol. 32, no. 4, pp. 46–54, 2015.
- [2] W. Zhao, D. Zhou, X. Qiu, and W. Jiang, "Compare the performance of the models in art classification," *PLOS ONE*, vol. 16, no. 3, e0248414, 2021.

- [3] W. Zhao, D. Zhou, X. Qiu, and W. Jiang, "How to represent paintings: A painting classification using artistic comments," *Sensors*, vol. 21, no. 6, 1940, 2021.
- [4] L. Cascone, M. Nappi, F. Narducci, and S. L. Russo, "Classification of fragments: recognition of artistic style," *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 4, pp. 4087–4097, 2023.
- [5] J. Kim, J. Y. Jun, M. Hong, H. Shim, and J. Ahn, "Classification of oil painting using machine learning with visualized depth information," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 42, pp. 617–623, 2019.
- [6] Y. Wang, Z. Xu, S. Bai, Q. Wang, Y. Chen, W. Li, X. Huang, and Y. Ge, "Intelligent painting identification based on image perception in multimedia enterprise," *Enterprise Information Systems*, vol. 16, nos. 10–11, pp. 1485–1499, 2022.
- [7] P. Abry, H. Wendt, and S. Jaffard, "When Van Gogh meets Mandelbrot: Multifractal classification of painting's texture," *Signal Processing*, vol. 93, no. 3, pp. 554–572, 2013.
- [8] S. T. Liong, Y. C. Huang, S. Li, Z. Huang, J. Ma, and Y. S. Gan, "Automatic traditional Chinese painting classification: A benchmarking analysis," *Computational Intelligence*, vol. 36, no. 3, pp. 1183–1199, 2020.
- [9] K. Georgoulaki, "Classification of Pointillist paintings using colour and texture features," *International Journal of Electrical and Computer Engineering Research*, vol. 2, no. 1, pp. 13–19, 2022.
- [10] S. Jiang, Q. Huang, Q. Ye, and W. Gao, "An effective method to detect and categorize digitized traditional Chinese paintings," *Pattern Recognition Letters*, vol. 27, no. 7, pp. 734–746, 2006.
- [11] C. Gao, J. Hou, and J. Gong, "Study on the symbiotic development path of traditional villages in the county based on the grey relation analysis: The case of Xunxian County in Henan, China," *Grey Systems: Theory and Application*, vol. 13, no. 1, pp. 193–205, 2023.
- [12] A. Amelio and C. Pizzuti, "An evolutionary approach for image segmentation," *Evolutionary Computation*, vol. 22, no. 4, pp. 525–557, 2014.
- [13] Y. Deng, F. Tang, W. Dong, F. Wu, O. Deussen, and C. Xu, "Selective clustering for representative paintings selection," *Multimedia Tools and Applications*, vol. 78, pp. 19305–19323, 2019.
- [14] J. Kim, J. Y. Jun, M. Hong, H. Shim, and J. Ahn, "Classification of oil painting using machine learning with visualized depth information," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 42, pp. 617–623, 2019.
- [15] J. Ao, Z. Ye, W. Li, and S. Ji, "Impressions of Guangzhou city in Qing dynasty export paintings in the context of trade economy: a color analysis of paintings based on k-means clustering algorithm," *Heritage Science*, vol. 12, no. 1, 77, 2024.
- [16] H. Yang and K. Min, "Classification of basic artistic media based on a deep convolutional approach," *The Visual Computer*, vol. 36, no. 3, pp. 559–578, 2020.
- [17] T. Chen and J. Yang, "A novel multi-feature fusion method in merging information of heterogeneous-view data for oil painting image feature extraction and recognition," *Frontiers in Neurorobotics*, vol. 15, 709043, 2021.
- [18] M. Wu, M. Li, and Q. Zhang, "Feature Separation and Fusion to Optimise the Migration Model of Mural Painting Style in Tombs," *Applied Sciences*, vol. 14, no. 7, 2784, 2024.
- [19] I. A. Albadarneh and A. Ahmad, "Machine learning based oil painting authentication and features extraction," *International Journal of Computer Science and Network Security (IJCSNS)*, vol. 17, no. 1, 8, 2017.
- [20] H. Jiang and T. Yang, "Research on the extraction method of painting style features based on convolutional neural network," *International Journal of Arts and Technology*, vol. 14, no. 1, pp. 40–55, 2022.
- [21] Z. Chen, A. Gu, X. Zhang, and Z. Zhang, "Authentication and inference of seal stamps on Chinese traditional painting by using multivariate classification and near-infrared spectroscopy," *Chemometrics and Intelligent Laboratory Systems*, vol. 171, pp. 226–233, 2017.
- [22] M. Aubry, B. C. Russell, and J. Sivic, "Painting-to-3D model alignment via discriminative visual elements," *ACM Transactions on Graphics (ToG)*, vol. 33, no. 2, pp. 1–14, 2014.
- [23] A. Sharma and K. K. Paliwal, "Linear discriminant analysis for the small sample size problem: an overview," *International Journal of Machine Learning and Cybernetics*, vol. 6, pp. 443–454, 2015.
- [24] T.-Y. Wu, H. Li, S. Kumari, and C.-M. Chen, "A Spectral Convolutional Neural Network Model Based on Adaptive Fick's Law for Hyperspectral Image Classification," *Computers, Materials & Continua*, vol. 79, no. 1, pp. 19–46, 2024.

- [25] F. Zhang, T.-Y. Wu, J.-S. Pan, G. Ding, and Z. Li, "Human motion recognition based on SVM in VR art media interaction environment," *Human-centric Computing and Information Sciences*, vol. 9, no. 40, 2019.
- [26] T.-Y. Wu, A. Shao, and J.-S. Pan, "CTOA: Toward a Chaotic-Based Tumbleweed Optimization Algorithm," *Mathematics*, vol. 11, no. 10, p. 2339, 2023.
- [27] X. Li, L. Yu, D. Chang, Z. Ma, and J. Cao, "Dual cross-entropy loss for small-sample fine-grained vehicle classification," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 5, pp. 4204–4212, 2019.
- [28] B. L. Menai and M. C. Babahenini, "The Effect Of Optimizers On CNN Architectures For Art Style Classification," *International Journal of Computing and Digital Systems*, vol. 13, no. 1, pp. 353–360, 2023.
- [29] J. Huang and C. X. Ling, "Using AUC and accuracy in evaluating learning algorithms," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 3, pp. 299–310, 2005.