MCCFNet: Multi-channel Color Fusion Network For Cognitive Classification of Traditional Chinese Paintings

Jing Geng[1] · Xin Zhang[1] · Yijun Yan[2] · Meijun Sun[3] · Huiyuan Zhang[1] · Maher Assaad[4] · Jinchang Ren[2] · Xiaoquan Li[2]

Abstract

The computational modeling and analysis of traditional Chinese painting rely heavily on cognitive classification based on visual perception. This approach is crucial for understanding and identifying artworks created by different artists. However, the effective integration of visual perception into artificial intelligence (AI) models remains largely unexplored. Additionally, the classification research of Chinese painting faces certain challenges, such as insufficient investigation into the specific characteristics of painting images for author classification and recognition. To address these issues, we propose a novel framework called multi-channel color fusion network (MCCFNet), which aims to extract visual features from diverse color perspectives. By considering multiple color channels, MCCFNet enhances the ability of AI models to capture intricate details and nuances present in Chinese painting. To improve the performance of the DenseNet model, we introduce a regional weighted pooling (RWP) strategy specifically designed for the DenseNet169 architecture. This strategy enhances the extraction of highly discriminative features. In our experimental evaluation, we comprehensively compared the performance of our proposed MCCFNet model against six state-of-the-art models. The comparison was conducted on a dataset consisting of 2436 TCP samples, derived from the works of 10 renowned Chinese artists. The evaluation metrics employed for performance assessment were Top-1 Accuracy and the area under the curve (AUC). The experimental results have shown that our proposed MCCFNet model significantly outperform all other benchmarking methods with the highest classification accuracy of 98.68%. Meanwhile, the classification accuracy of any deep learning models on TCP can be much improved when adopting our proposed framework.

Keywords Visual cognition · Multi-channel color fusion network (MCCFNet) · Regional weighted pooling (RWP) · Chinese painting classification

Introduction

Chinese painting, referred to as "Traditional Chinese Painting (TCP)," is a traditional form of painting in China spanning thousands of years. Being painted usually on silk or rice paper with a brush dipped in water, ink, or painting pigment, TCP reflects the artists' understanding of nature, society and the related politics, philosophy, religion, morality, literature, and art.

By using the blank-leaving way in painting, TCP is featured by the artistic effect of the "coexistence of the virtual and the real." Furthermore, TCP values the combination of calligraphy and poetry, decorating with seals, emphasizing the connections between the art and the nature with the concept of "writing the spirit with shape". From the perspective of techniques, TCP can be categorized into "Gongbi,"

✉ Yijun Yan
  y.yan2@rgu.ac.uk

✉ Jinchang Ren
  j.ren@rgu.ac.uk

[1] Faculty of Printing, Packaging Engineering and Digital Media Technology, Xi'an University of Technology, Xi'an 710048, China

[2] National Subsea Centre, Robert Gordon University, Aberdeen AB21 0BH, UK

[3] College of Intelligence and Computing, Tianjin University, Tianjin, China

[4] Department of Electrical and Computer Engineering, Ajman University, P.O. Box 346, Ajman, United Arab Emirates

"Xieyi," and "Baimiao." The 'Xieyi" technique is featured by exaggerated forms and freehand brushwork. On the other hand, the brushwork in "Gongbi" paintings is fine and visually complex [1]. The "Baimiao" is one of the steps of Chinese "Gongbi" painting, which is a classic ink and brush line drawing, and is referred to a linear art due to its concise and monochrome attributes. It will become a Chinese "Gongbi" painting if we color it; alternatively, it can be an independent art style. Figure 1 shows specific examples of the painting images used in this paper, including the genre of each painting and its corresponding techniques and painters.

As a cultural heritage of China and even the whole world, TCPs have eternal value due to their inimitability, especially those extremely precious ones that have been handed down for thousands of years. Specifically, these should be mounted and stored in a place with extremely low humidity, no sunlight, and dust. For some reasons, special TCPs cannot be mounted, and their storage needs to be particularly extra careful without any folding or stacking. Otherwise, break-off and black seal at the fold part will occur, causing permanent and unrepairable damages. Due to these restrictive requirements, the way to the preservation of TCPs usually takes much manual labor and resources. To foster the preservation effectiveness and the efficiency of archiving and retrieving the TCPs as well as reduce the staff workload and prevent unexpected damages from manual operations, it is crucial to explore the artificial intelligence (AI) driven computational analysis of TCPs for automated classification. Due to its rich expressiveness, characterizing TCPs usually rely on cognition and visual perception [2], which has inspired our proposed multi-channel color fusion framework (MCCFNet), aiming at exploiting the global attributes of the TCP. Given an input color image of a TCP, we will first convert it into multiple color spaces. Each color space map will be fed into an improved DenseNet169 where the final layer will be extracted as the global feature vector. Finally, multiple feature vectors will be concatenated together followed by a classifier, i.e., support vector machine (SVM), for classification. In the experiment, we compared our approach with 6 benchmarking models. The experimental results have fully validated the effectiveness and efficiency of our method.

Fig. 1 The theme, painting techniques and painters of Chinese painting

The main contributions and the innovation of this paper can be summarized as follows.

1. Inspired by visual cognition, we propose a multi-channel color fusion framework (MCCFNet) for the classification of TCPs.
2. Taking the DenseNet169 as the backbone, we introduce a multiple color channel fusion module and a new Regional weighted pooling (RWP) to improve the feature effectiveness in MCCFNet. RWP can overcome average pooling caused feature loss and maximum pooling caused gradient spread faults.

The remainder of the paper is organized as follows. The "Related Works" section describes the related work of TCP classification. The "Proposed MCCFNet Model" section details the proposed MCCFNet. The "Experimental Results and Discussion" section presents the experimental results and corresponding discussions. Finally, some conclusions and future directions are summarized in the "Conclusions" section.

Related Work

There are two commonly used features for the classification of painting images, including color features and texture features [3]. Color is the most intuitive factor of representation in paintings. In [4], color moments, hue and color range information are used to identify the subject matter of landscapes, flowers and birds, and people, as well as the authors of TCP. In [5], superpixels are generated for segmenting the region of interest, followed by the convolutional neural network (CNN) to extract semantic features before classification.

Texture is a visual attribute that captures the surface structure and layout of objects with gradual or periodic changes, representing homogeneous patterns in images. Unlike pixel-based color features, texture features rely on region based statistics. Consequently, texture alone may not effectively represent the intrinsic characteristics of an image, limiting its contribution to classification based decision-making. To validate this, we compared the efficacy of color and texture

features using a deep learning model on our own dataset, i.e., three color feature domains (RGB, HSV, L*a*b*) and one texture domain (gray co-occurrence matrix), as presented in Table 1. An evident drawback of texture features is their sensitivity to the changes in image resolution, which can lead to significant deviations in calculated texture values.

In contrast, color characteristics offer distinct advantages. Firstly, each Chinese painter exhibits their own unique style. For instance, Qi Baishi's paintings are characterized by rich and vibrant colors, highlighting their prominent color characteristics. Pan Tianshou's works feature intricate interplay between ink and color, resulting in a dim color palette. Zeng Xiaolian's paintings primarily focus on the physical characteristics of objects, showcasing relatively simple and elegant colors. Secondly, in Chinese painting images, the background typically consists of off-white paper, while the painted areas depict the artist's interpretation of object colors. This leads to noticeable regional color differences in the global color features. Compared to texture features, color characteristics are more effective in expressing the most representative attributes of the image itself.

In [6], edge detection from the Sobel operator and morphology processing is applied to connect the stroke features of paintings. In [7], a multi-task joint sparse representation algorithm is proposed for extracting texture features before the classification of TCPs. In [8], an end-to-end multi-task feature fusion method is proposed to fuse the semantic features and texture features. Gram matrix [9] is used to represent painting styles, where the style transfer work has achieved good results. It has also been used to represent painting techniques and stroke styles [10]. In addition, frequency domain decomposition is also used to extract texture features, such as using the discrete cosine transform (DCT) and edge features to identify "Gongbi" and "Xieyi" [11].

With the wide applications of deep learning, automatic feature extraction and classification are implemented in parallel in an end-to-end manner [12]. Starting from AlexNet [13], a series of improved models are introduced, such as VGGNet [14], GoogLeNet [15], Inception V3 [16], and DenseNet [17]. In [18], the performance of different deep networks are compared using the same dataset of painting images. By taking style classification as an example, as shown in Table 2, Inception V3, ResNet [19] and DenseNet

Table 1 The accuracy of different networks in different spaces (accuracy)

Feature domain	Models					
	DenseNet169 [17]	MobileNetV2 [38]	ResNet50 [19]	VGG16 [14]	VGG19 [14]	Xception [39]
RGB	0.851	0.777	0.859	0.761	0.647	0.869
HSV	0.839	0.746	0.819	0.703	0.5989	0.685
L*a*b*	0.829	0.812	0.816	0.717	0.6374	0.867
GLCM	0.639	0.582	0.637	0.503	0.5519	0.650

Table 2 Accuracy of painting style classification from different deep neural networks (%), where the results are adopted from [21–24] and V* and R# denote VGG and ResNet, respectively

Literature	Dataset	Data/category	AlexNet	GoogleNet	InceptionV3	V*-13	R#-50	DenseNet
[21]	Self-built	797/17	-	69.90	79.26	-	76.48	79.36
[22]	WikiArt	30,870/6	62.46	64.42	67.16	-	66.64	-
[23]	WikiArt	80,000/25	37.80	-	-	-	49.40	-
[24]	WikiArt	81,449/20	58.20	-	-	60.10	-	-

have produced better results than other networks, due mainly to the residual connection used.

In addition, optimization of the deep learning network can further improve the performance, such as [20], where the feature layer of VGG16 network is cut to constructed the new structure of VGG15, leading to reduced error rate of TCP subject recognition by 8.8%.

Although the dataset configurations are not exactly the same, some interesting observations can be found from Table 2. For the large volume of the WikiArt dataset, deep learning-based automatic feature extraction is widely used, because of their flexible feature extraction capabilities. As the amount of data increases, there are fewer manual-crafted features used, yet more expert knowledge is added in the form of fused features to improve the classification accuracy. Accuracy of painting style classification from different deep neural networks (%), where the results are adopted from [21–24], V* and R# denote VGG and ResNet, respectively.

In addition, Jiang et al. [1] proposed a classification model combining the discrete cosine transform and CNN. Eli David et al. [25] proposed to use deep convolutional self-coding neural network for classification of TCPs even with a small number of samples, which reduced the error rate by 63% compared with the traditional methods in the classification of three painters. Lecoutre et al. [26] integrated a pre-trained deep neural network with residual units to classify the representation of western paintings, and the performance of the algorithm was verified on the Wiki painting dataset.

For feature extraction from depth learning, CNN has become the main mainstream method, which has been successfully applied in image and video recognition, recommendation system and natural language processing [27]. With the further development of the ResNet, it establishes a "shortcut" (skip connection) between the front layer and the back layer, which helps to reverse the gradient in the training process and enable the training of a deeper CNN. Another popular model is DenseNet, which is composed of several dense blocks where a dense connection between all the previous layers and the subsequent layers is established. By reusing features through the dense connections, it enables the DenseNet to achieve better performance than the ResNet with fewer parameters and computing costs.

Therefore, in our experiment, we used the DenseNet as the backbone network.

For human beings, the identification of artistic paintings is very subjective. Although existing models can extract features from the artworks, they rarely consider and analyze the perspective of human visual perception. Therefore, how to integrate visual cognition into the recognition and classification of different painting styles is still an undeveloped field. It is our aim to tackle with this particular challenge, where the proposed model will be detailed in the next section.

Proposed MCCFNet Model

Color is one of the most significant features of human visual perception. In previous studies, the expression of color features in painting images was not prominent enough. In response to this research pain point, we propose a multi-channel color feature fusion model; In order to compensate for image blur caused by average pooling and feature loss caused by maximum pooling, this study proposes a new weighted pooling algorithm (RWP).

As shown in Fig. 2, the color feature fusion model for this channel (MCCFNET) composed of five main stages, i.e., color space transformation, pre-processing, weighted pooling (RWP), feature extraction, and classification-based decision-making. The images of three color channels are used as input data, which are respectively imported into the residual network DenseNet169 improved by RWP. They are output as feature vectors in the fully connected layer and then fused through feature fusion before being input into the support vector machine SVM for final classification.

Color Space Transformation

Although red, green, and blue (RGB) is the most commonly used color space in life, its disadvantage is the potentially high correlation of the three channels [28]. As the sensitivity of human eyes to the three colors is different, the uniformity of RGB color space is very poor. Hue, saturation, and value (HSV) is a relational representation of color in an alternative color space, which aims to describe a more accurate color perception than RGB, yet remains computationally
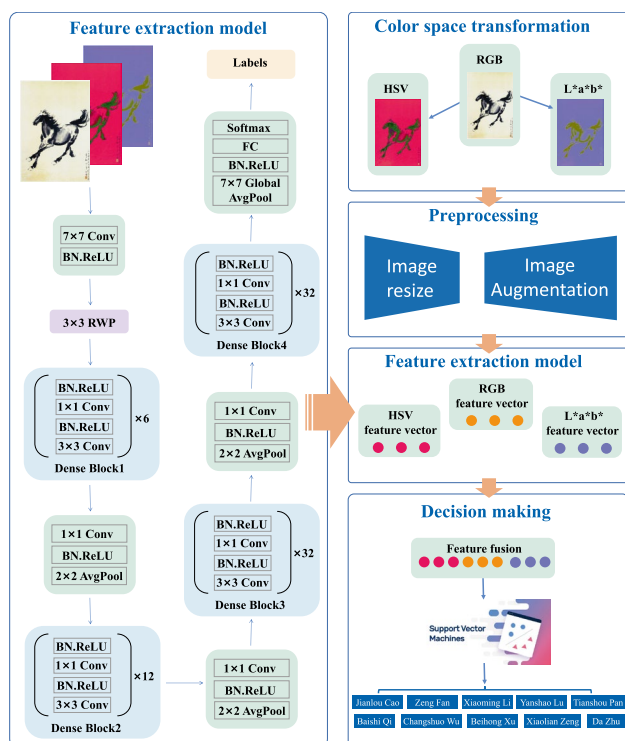
Fig. 2 The architecture of the proposed MCCFNet

Fig. 3 Forward propagation of the forward RWP

efficient. In addition, L*a*b* is another color space, which is designed to mitigate the human vision, focusing mainly on the perception uniformity, where the L* component closely matches human brightness perception. Therefore, it can be used to make accurate color balance by modifying the output color scale of a* and b* components or to adjust the brightness contrast by modifying L* components. These transformations are difficult or impossible in the RGB space as it is modeled on the output of physical devices, rather than the human visual perception.

Inspired by above-mentioned visual perception, in our MCCFNet, we mainly select RGB, HSV and L*a*b* color spaces. Three color spaces will be fed into three identical deep neural network branches for extraction of deep color features, as they supplement to each other as discussed above. The final layer of each branch, with a size of 10,241, will be concatenated followed by decision-making using the support vector machine (SVM) as the classifier.

Weighted Pooling

For conventional pooling, the weights in the pooling window are set to be the same. However, when identifying texture and edge information, which can only be distinguished by detailed information, their disadvantages will be further
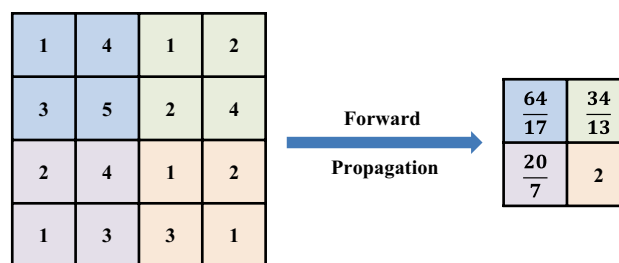
amplified. To solve this problem, a 3×3 regional weighted pooling (RWP) layer is proposed to strengthen the ability of feature extraction.

Maximizing pooling can obtain local information and better preserve the texture features, but it may result in serious information loss due to the discarded the non-maximum values. The average pooling, however, can often retain the characteristics of the overall data and highlight the background information while blurring the images. In addition, the error of feature extraction of CNN comes mainly from two aspects [13]. On the one hand, the variance of estimated value increases due to the limitation of neighborhood size. On the other hand, the deviation of the estimated mean value is increased due to the parameter error of the convolution layer [29]. To solve this problem, a weight pooling layer which fuses the maximum pooling rule and average pooling rule was proposed, where the back propagation processing is missed. Therefore, we propose a whole procedure of RWP, for the forward propagation process of regional weight proportional pooling, given an input data with the size of M×N, the size of pooling window is $n \times n$, the value in the pooling window region is $a_1, a_2, \cdots a_{n \times n}$, the weight ratio of each value is $w_1, w_2, \cdots w_{n \times n}$, the pooled result r of the pooled window region can be calculated according to Eqs. (1–2). Figure 3 shows a specific example of the forward propagation of the forward RWP.

$$r = \sum_{i=1}^{n \times n} w_i * a_i \tag{1}$$

$$w_i = \frac{a_i + |a_{\min}|}{\sum_{j=1}^{n \times n}(a_j + |a_{min}|)} \tag{2}$$

The backpropagation is defined in Eq. (3), where the new value b_i in the pooling window region after upsampling can be calculated. Figure 4 shows a specific example of the back propagation of the backward RWP.

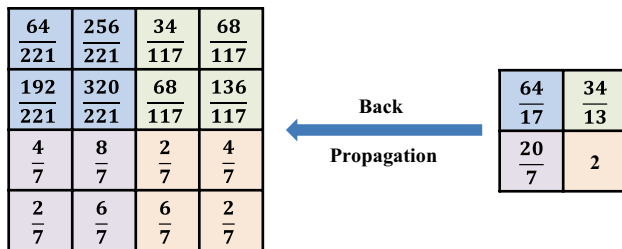$$b_i = \frac{a_i}{\sum_{i=1}^{n \times n} a_i} * r \tag{3}$$

$\frac{64}{221}$	$\frac{256}{221}$	$\frac{34}{117}$	$\frac{68}{117}$
$\frac{192}{221}$	$\frac{320}{221}$	$\frac{68}{117}$	$\frac{136}{117}$
$\frac{4}{7}$	$\frac{8}{7}$	$\frac{2}{7}$	$\frac{4}{7}$
$\frac{2}{7}$	$\frac{6}{7}$	$\frac{6}{7}$	$\frac{2}{7}$

Back **Propagation** →

$\frac{64}{17}$	$\frac{34}{13}$
$\frac{20}{7}$	2

Fig. 4 Back propagation of the backward RWP

Data Processing

Preprocessing

Prior to training, the images are resized to a standardized dimension, i.e., 224×224, to ensure consistency across the dataset. Rescaling the images helps in maintaining spatial information and facilitating compatibility with deep neural network architectures. Additionally, various data augmentation techniques are applied to increase the diversity and quantity of the training data. These techniques, including random rotation, flipping, clipping, scaling, and shifting, introduce variations to the images without altering their semantic content, leading to enhanced performance and improved accuracy. By augmenting the data, the overfitting problem can be mitigated, while the model's generalization ability is also improved.

Feature Extraction

As stated in [21–24], DenseNet [17] is more efficient than other networks, where the key lies in the reduction of the computational cost due to the reused features. More importantly, DenseNet can exploit high-level semantic information and alleviate the problem of gradient disappearance. To this end, it is employed as the backbone network in our MCCFNet.

The weighted pooling strategy simulates diverse perceptions of different image components. By integrating the weighted pooling strategy, the representative features can be extracted, where the RWP can well retain the important information of the feature map and improve the accuracy of the model while maintaining efficient computing and memory as validated in the experimental results.

Decision-Making

Taken the last fully connected layer as the feature vector, three feature vectors from three color spaces will be fused together by concatenation, followed by using the SVM classifier for decision-making. In our framework, we use the

SVM with a radial basis function (RBF) kernel as it has been widely used for painting style recognition and is also considered to be suitable for classification of TCPs [30].

Experimental Results and Discussion

Experimental Setting

In our experiments, we gathered a collection of 2436 paintings by 10 renowned Chinese artists from esteemed institutions such as the Palace Museum and Tianjin Museum. The dataset includes works by various artists, each with their own distinct style and focus. Specifically, we have 318 paintings by Jianlou Cao, known for their expertise in freehand brushwork of flowers, vegetables, and fruits, with a particular emphasis on pomegranates. Zeng Fan contributed 91 works, showcasing exceptional sketching skills and a penchant for freehand characters. Xiaoming Li's 92 paintings specialize in meticulous flower and bird depictions, while Yanshao Lu presents 236 works primarily focusing on landscapes. Tianshou Pan, on the other hand, offers 83 works showcasing expertise in freehand brushwork of flowers, birds, and landscapes, occasionally featuring characters. Baishi Qi, a highly prolific artist, contributed 904 works encompassing a wide range of subjects, including flowers and birds, insects and fish, landscapes, and characters. Changshuo Wu's collection of 135 works excels in freehand brushwork of flowers and occasionally incorporates landscapes. Beihong Xu is known for their 192 works that showcase exceptional skill in depicting characters, animals, flowers, and birds and are particularly renowned for their dynamic horse paintings. Xiaolian Zeng's 240 works are highly regarded for their realistic depiction of plants, flowers, animals, and birds. Lastly, Da Zhu's 145 works captivate viewers with exquisite ink freehand brushwork, especially in flower and bird paintings.

For a fair evaluation, we divided the images into non-overlapping training, validation, and testing sets in proportions of 70%, 10%, and 20%, respectively. All experiments were conducted on a RTX2080Ti graphics processing unit (GPU). For our proposed MCCFNet, the core lies in an improved DenseNet-169 network, comprising four Dense-Blocks with varying numbers of dense connections: 6, 12, 32, and 32, respectively. Initially, a 7×7 convolution layer is employed to downscale the image from 224×224 pixels to 112×112 pixels. This is followed by the application of a batch normalization (BN) layer and a rectified linear unit (ReLU) layer for feature enhancement. In the training process, the optimizer used is stochastic gradient descent [31], which facilitates faster convergence. The momentum value is set to 0.9, while the learning rate is set to 0.003. The activation function selected for the model is ReLU, and dropout

Table 3 Comparison of the classification accuracy from different approaches

Models	Top-1 Accuracy (%)
Saleh and Elgammal [32]	63.06
Tan et al. [33]	76.11
Huang et al. [34]	81.87
Sheng J C et al. [35]	83.32
Li [36]	74.17
Jiang [37]	94.93
MCCFNet	**98.68**

regularization with a rate of 0.5 is incorporated. The batch size is set to 16. For all benchmarking methods, the default settings are adopted for consistency.

Results and Discussion

For objective evaluation, several quantitative metrics are used, which include the Top-1 Accuracy and the AUC.

The probability that the score of a positive class sample is greater than that of a negative class sample for any given positive class sample and a negative class sample.

The formula is as follows: M is the number of positive samples, and N is the number of secondary samples:

$$AUC = \frac{\sum_{i \in positive} rank_i - \frac{M*(1+M)}{2}}{M * N} \quad (4)$$

Comparison Against Other Methods

To validate the effectiveness of our proposed model, several state-of-the-art deep learning models were selected for comparison in Table 3.

- Saleh and Elgammal [32] integrated visual features and metric learning together to learn the optimized similarity between paintings, followed by a metric learning model, Feature fusion and metric-fusion for semantic-level decision such as predicting a painting's style, genre, and artist.

- Tan et al. [33] proposed a CNN-based model for fine-art painting classification.
- Huang et al. [34] proposed a novel two-channel (including the RGB channel and the brush stroke information channel) deep residual network to classify fine-art painting images.
- In Sheng et al. [35], a wavelet transform-based model was proposed to extract the texture feature of paintings followed by three machine learning methods (decision tree, neural network, and SVM) for classification.
- Li [36] proposed a statistical framework for assessing the averaged distinction level of paintings for classification.
- Jiang [37] proposed an end-to-end multi-task architecture to extract both color and texture features and classify traditional Chinese paintings.

Most of the above methods use depth learning and AI-driven machine learning models to extract the features from the paintings and make the classification. Due to lack of cognitive perception, the performance of these models is still limited. Thanks to our cognitive framework, the way our proposed method classifies the paintings fits human visual perception very well, which helps our proposed MCCFNet model significantly outperform all other benchmarking methods with the highest classification accuracy of 98.68%.

Comparison of Different Backbone Models

In this section, we evaluate the classification performance of different backbone models. Table 4 presents the Top-1 Accuracy and AUC produced by embedding various models such as DenseNet-169 [17], ResNet-50 [19], VGG16 [14], VGG19 [14], MobileNetV2 [38], and Xeption [39] to our proposed MCCFNet. Except for the ResNet50, all other mainstream models have more than 95% Top-1 Accuracy which has surpassed the best benchmarking method Jiang [37], indicating the effectiveness of our proposed framework. Table 4 also presents the AUC value of 7 different backbone models, where the Improved DenseNet169 for the ten categories of paintings exceeds 0.999, which again validates the effectiveness of the improved DenseNet169 we proposed.

Table 4 Comparison of different backbone models in MCCFNet

Metrics	Network						
	Improved DenseNet169	DenseNet169 [17]	MobileNetV2 [38]	ResNet50 [19]	VGG16 [14]	VGG19 [14]	Xception [39]
Top-1 Accuracy	**98.68**	97.34	95.40	93.12	97.26	98.05	96.45
AUC	**99.91**	99.86	99.66	97.26	98.21	99.73	99.78

Table 5 RWP ablation experiments

Backbone model	Pooling strategy		
	Max pooling [42]	Average pooling [41]	**RWP**
VGG16	92.05	90.40	**92.75**
ResNet50	96.44	94.00	**97.02**
DenseNet169	93.87	93.54	**94.04**

Ablation Experiment

The error in feature extraction of the convolutional neural networks (CNNs) can be attributed to two main factors [13]. Firstly, the limited neighborhood size of convolutional operations increases the variance of estimated values [40]. Secondly, errors in the parameters of the convolutional layers result in a shift in the estimated mean. To address these issues, different pooling strategies are employed.

In the context of Chinese painters' brush and ink techniques, the application of average pooling in DenseNet [41] effectively reduces the first error but can lead to the loss of crucial texture information. Conversely, max pooling [42] reduces the second error but sacrifices domain feature information. To strike a balance and improve the effectiveness of feature extraction in MCCFNet, we utilize a regional weight pooling (RWP) rule. This rule combines the maximum pooling and average pooling strategies to minimize both types of errors and optimize the feature extraction process.

For backpropagation rules, both average pooling and maximum pooling focus on pixels that contribute significantly to features during forward propagation. However, average pooling backpropagation introduces errors between the propagated gradient and its true value for each pixel. On the other hand, maximum pooling backpropagation only propagates the gradient to a single point, disregarding the contribution of other points to classification decisions. In this case, RWP can address these issues to ensure the accuracy of each point's contribution to the features.

To evaluate the effectiveness of weighted pooling, we conducted RWP ablation experiments using popular neural network frameworks such as VGG16, ResNet50, and DenseNet169. We compared the results with commonly used maximum pooling and average pooling methods by replacing the pooling rules in the original framework with corresponding comparison pooling rules. We then classified our Chinese painting dataset.

The results, as shown in Table 5, demonstrate the superiority of RWP over maximum pooling and average pooling in the three neural network frameworks. RWP achieved an approximately 0.5 percentage point advantage. This weighted pooling operation assigns different weights based on the importance of different positions, resulting in improved accuracy and efficiency of pooling operations.

Comparison With Different Color Spaces

In Table 6, we further compare the performance of different combinations of color spaces in various backbone models for feature extraction in our framework. As seen, for each backbone model, multiple color space inputs can achieve better classification accuracy than using a single color space, which has validated the effectiveness of the concept of multi-channel color space fusion. On the other hand, with a single color space as input, there is no obvious winner among the three color spaces used, which indicates their supplementary nature hence the fused strategy. With double input of color spaces, RGB + HSV and RGB + L*a*b* can always perform better than HSV + L*a*b*, possibly because RGB color space is emphasized in the pre-trained deep learning models. The results from the last column of the Table 4 show that the accuracy of improved DenseNet169 can surpass all other benchmarking models when fusing all the three color spaces of RGB, HSV, and L*a*b*. Again, this has validated the efficacy of our improved DenseNet169, as well as the necessity of color space fusion. However, due to the fact of inherent limitation, some models (i.e., MobileNetV2 and Xception) produce inferior results with 3 color spaces than

Table 6 Comparing the Top-1 accuracy (%) of different models in various combination of color spaces

Model	Color space						
	HSV	L*a*b*	RGB	HSV + L*a*b*	RGB + HSV	RGB + L*a*b*	RGB + HSV + L*a*b*
DenseNet169 [17]	93.54	91.06	92.88	94.87	97.85	97.85	98.01
MobileNetV2 [38]	84.60	85.60	87.58	89.40	96.19	96.19	96.19
ResNet50 [19]	87.25	89.90	85.10	90.73	94.70	94.54	94.87
VGG16 [14]	88.08	92.55	92.05	92.72	96.36	97.19	98.01
VGG19 [14]	91.39	89.90	92.88	93.54	98.01	97.52	98.34
Xception [39]	91.23	90.56	92.88	93.54	97.35	98.01	97.68
Improved DenseNet169	**92.22**	**92.55**	**93.54**	**95.03**	**98.51**	**98.01**	**98.68**

those using only two color spaces. This is probably because both these lightweight models fail to extract the as discriminative features as other complexed models do.

Classification Error Analysis

Figure 5 shows the confusion matrices of the classification results in three different backbone models. Taking the improved DenseNet169 as an example, we will analyze it and interpret the results from the perspective of art. As seen, there is 5% of misclassification from Pan to Fan, which can be possibly explained as follows. As a master of "Xieyi" style, Fan often imitates Zhu's landscape works where the line drawing technique is often used with the layers and stripes in the painting. Pan is also a master of "Xieyi" style, focusing mainly on drawing flowers, birds, and landscapes. As a student of Zhu, Pan's paintings not only inherit Zhu's atmospheric and elegant style in painting, but also feature his own strong and forceful characteristics. Due to the fact that Fan and Pan are two masters from the same school, it is not surprising that their works are somehow similar to each other. In addition, about 6% of paintings from Wu was wrongly classified to Cao. According to the historical survey, Cao is the best third generation descendant of Wu. Therefore, both of them are the masters of drawing flowers, fruits, vegetables, et al., and their strokes are smooth and vigorous. It is interesting to find that the misclassification caused by AI models seems to be consistent with the real art heritage.

Figure 6a shows the misclassified images during the classification process. Among the 10 painters' paintings, six of them were misclassified. Moreover, among these misclassified cases, it is interested to find that the two paintings of Wu are misclassified as Cao's paintings, as seen in the paintings shown in the first row of Fig. 6b. For a subjective comparison, we also show two Cao's paintings in the second row of Fig. 6b.

The first misclassified painting fully reflects the diagonal composition technique of Wu's flower painting, which is one of the four artistic features. His plum blossom, orchid, etc., regardless of banners or vertical paintings, mostly rise obliquely from the bottom left to the top right, forming a trend of oblique rise. Most lines are zigzag and like to be left blank in the painting. Cao also has similar painting techniques as Wu, which can be clearly seen from his paintings.

The second feature of Wu's painting is the very thick ink used, and he is good at using solid colors such as red and green. As seen from the painting examples, the rich colors in this painting are not vulgar, but show a feeling of both refined and popular appreciation. Cao's paintings are also found to be very eye-catching and strong in color, with thick ink and light color.

Another technique of Wu's flower painting, as one of the four major artistic features, is the usage of the power of the brush to penetrate the back of the paper and fully apply the experience of calligraphy and seal cutting to painting creation. As both Wu and Cao like to paint forcefully, Wu's paintings are not surprisingly misclassified as Cao's.

Wu also likes poetry. In his paintings, Wu's poems are organically integrated within the pictures. In Cao's paintings, he often uses poetry to express his thoughts and feelings as well. From the above facts, it explains why AI models have mistakenly classified Wu's painting as Cao's, which is consistent with the law of painting to some extent.
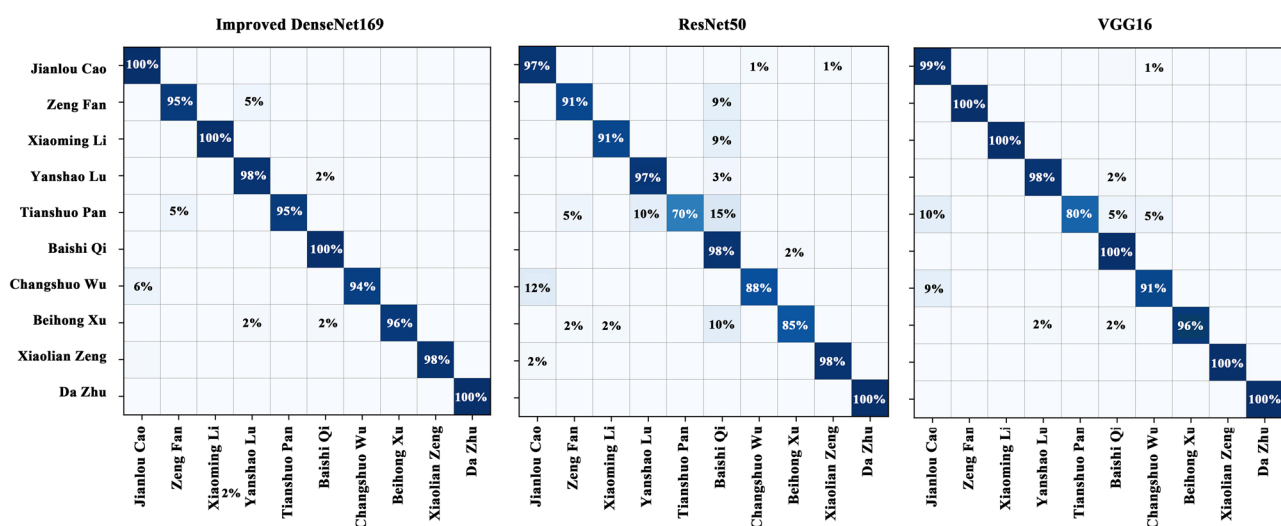

Fig. 5 Visual comparison of the confusion matrices for TCP classification in different backbones
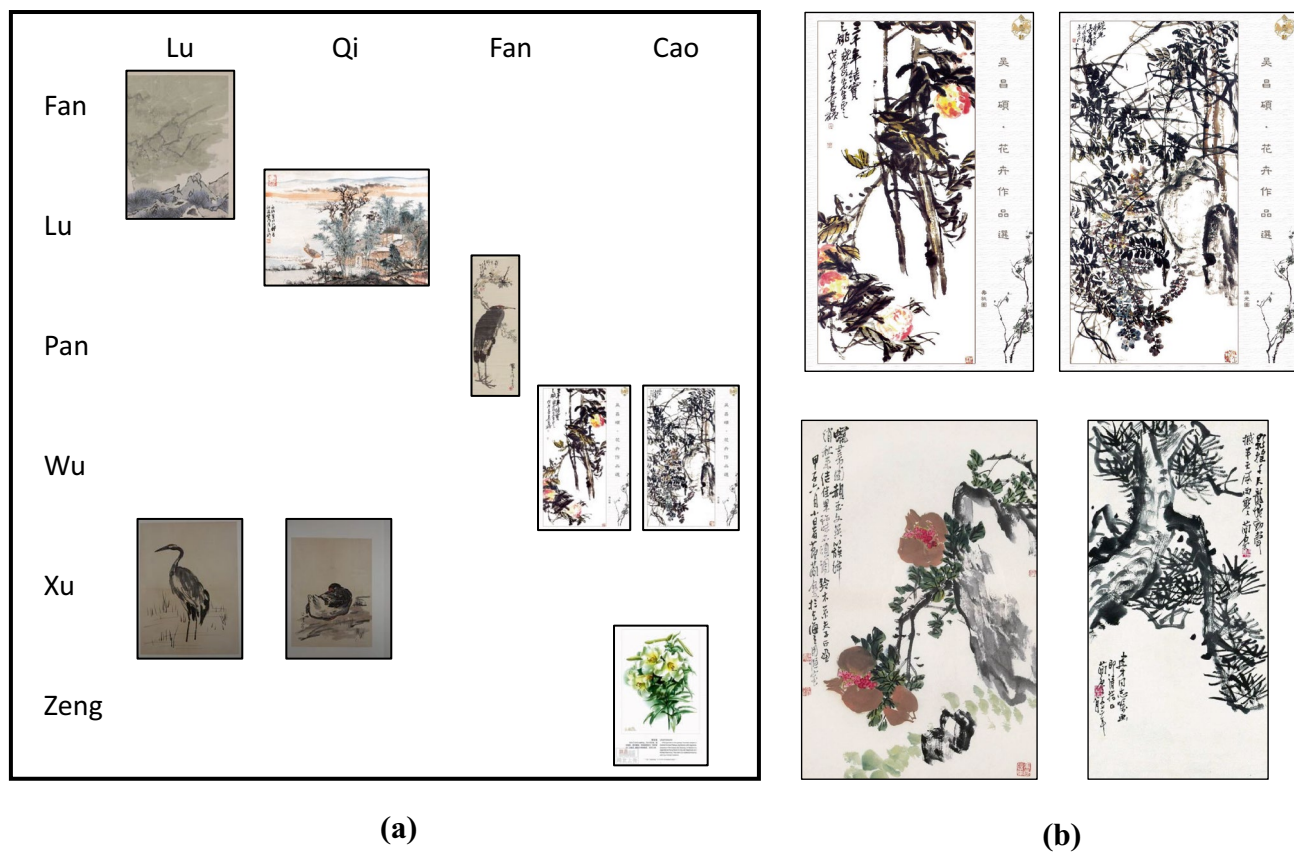
Fig. 6 a Illustration of misclassification cases, **b** wrongly divided into Cao's painting (top) and Cao's own painting (bottom)

Conclusions

In this paper, we introduce MCCFNet, a novel color feature extraction framework inspired by visual cognition, specifically designed for characterizing Chinese paintings. Our framework leverages an enhanced DenseNet169 model to extract features from three different color spaces. These features are then fused at the final layers of the branches and passed to an SVM classifier for classification. Experimental results demonstrate that MCCFNet achieves superior precision, recall, and F1 measures compared to several state-of-the-art deep learning models. Importantly, our study validates the concept of multi-channel color fusion, as incorporating multiple color spaces as input significantly improves the classification accuracy of Chinese paintings. Additionally, our proposed regional weighted pooling strategy enhances the performance of DenseNet169, surpassing other backbone models utilized in our experiments.

To further enhance the classification accuracy and reduce the computational costs, we plan to explore the integration of saliency detection within our model. This technique can effectively highlight the target within the image by distinguishing it from the background [43].

Moreover, we aim to extract and combine emotional features from Chinese paintings, as they provide rich insights into the feelings and expressions of the artists. To achieve this, we will delve into the realm of aesthetic and cognitive psychology to gain a deeper understanding of the psychological aspects related to the interpretation of these artworks. We will explore state-of-the-art networks such as MobileNet [44] and deep residual network [45] to facilitate the extraction and integration of emotional features, further enriching the classification process.

Data Availability The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Ethics Approval This article does not contain any studies with human participants performed by any of the authors.

Conflict of Interest The authors declare no competing interests.

References

1. Jiang W, Wang Z, Jin JS, Han Y, Sun M. DCT-CNN-based classification method for the Gongbi and Xieyi techniques of Chinese ink-wash paintings. Neurocomputing. 2019;330:280–6.

2. Wu CQ. Neural substrates and temporal characteristics for consciousness, brief sensory memory, and short-term memory (STM) systems. Proceedingsof the Annual Meeting of the Cognitive Science Society. 2005;27(27):2577.

3. Li J, Wang JZ. Studying digital imagery of ancient paintings by mixtures of stochastic models. IEEE Trans Image Process. 2004;13(3):340–53.

4. Sun M, Zhang D, Wang Z, Ren J, Jin JS. Monte Carlo convex hull model for classification of traditional Chinese paintings. Neurocomputing. 2016;171:788–97.

5. Sheng J, Jiang J. Style-based classification of Chinese ink and wash paintings. Opt Eng. 2013;52(9):093101–1 - 093101–8.

6. Wang Z, Sun M, Han Y, Zhang D. Supervised heterogeneous sparse feature selection for Chinese paintings classification. J Comput Aided Des Comput Graph. 2013;25(12):1848–55.

7. Yin XC, Yin X, Huang K, Hao HW. Robust text detection in natural scene images. IEEE Trans Pattern Anal Mach Intell. 2014;36(5):970–83.

8. Gupta S, Arbeláez P, Girshick R, Malik J. Indoor scene understanding with RGB-D images: bottom-up segmentation, object detection and semantic segmentation. Int J Comput Vision. 2015;112(2):133–49.

9. Bao H, Liang Y, Liu HZ, Xu D. A novel algorithm for extraction of the scripts part in traditional Chinese painting images. In 2010 2nd International Conference on Software Technology and Engineering. IEEE, 2010;2:V2–26-V2–30.

10. Jiang S, Huang Q, Ye Q, Gao W. An effective method to detect and categorize digitized traditional Chinese paintings. Pattern Recogn Lett. 2006;27(7):734–46.

11. Sheng JC, An effective approach to identify digitized IWPs (ink and wash paintings). 5th International Congress on Image and Signal Processing. IEEE. 2012;2012:407–10.

12. Li Y, Ren J, Yan Y, Liu Q, Ma P, Petrovski A, et al. CBANet: an end-to-end cross band 2-D attention network for hyperspectral change detection in remote sensing. IEEE Transactions on Geoscience and Remote Sensing. 2023; Early Access.

13. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun ACM. 2017;60(6):84–90.

14. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv.1409.1556. 2014;1–14.

15. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. Proceeding of the IEEE conference on computer vision and pattern recognition. 2015; 1–9.

16. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016;2818–2826.

17. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017;4700–4708.

18. Yue Lu, Chao G, Yi-Lun L, Fan Z, Fei-Yue W. Computational aesthetics of fine art paintings: The state of the art and outlook. Acta Automatica Sinica. 2020;46(11):2239–59.

19. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016;770–778.

20. Meng Q, Zhang H, Zhou M, Zhao S, Zhou P. The classification of traditional Chinese painting based on CNN. Cloud Computing and Security: 4th International Conference. 2018;232–241.

21. Kelek MO, Calik N, Yildirim T. Painter classification over the novel art painting data set via the latest deep neural networks. Procedia Computer Science. 2019;154:369–76.

22. Sandoval C, Pirogova E, Lech M. Two-stage deep learning approach to the classification of fine-art paintings. IEEE Access. 2019;7:41770–81.

23. Lecoutre A, Negrevergne B, Yger F. Recognizing Art Style Automatically with deep learning. Asian conference on machine learning. PMLR, 2017;327–342.

24. Elgammal A, Liu B, Kim D, Elhoseiny M, Mazzone M. The shape of art history in the eyes of the machine. Proceedings of the AAAI Conference on Artificial Intelligence. 2018;32(1):2183–91.

25. David OE, Netanyahu NS. DeepPainter: painter classification using deep convolutional autoencoders. Artificial Neural Networks and Machine Learning–ICANN 2016: 25th International Conference on Artificial Neural Networks. 2016;20–28.

26. Bay H, Tuytelaars T, Van Gool L. Surf: Speeded up robust features. Lect Notes Comput Sci. 2006;3951:404–17.

27. Shao L, Zhu F, Li X. Transfer learning for visual categorization: a survey. IEEE transactions on neural networks and learning systems. 2014;26(5):1019–34.

28. Halawani A, Burkhardt H. On using histograms of local invariant features for image retrieval. MVA. 2005;538–541.

29. Dumoulin V, Visin F. A guide to convolution arithmetic for deep learning. arXiv preprint arXiv:1603.07285. 2016;1–31.

30. Niazmardi S, Demir B, Bruzzone L, Safari A, Homayouni S. Multiple kernel learning for remote sensing image classification. IEEE Trans Geosci Remote Sens. 2017;56(3):1425–43.

31. Sutskever I, Martens J, Dahl G, Hinton G. On the importance of initialization and momentum in deep learning. International conference on machine learning. PMLR, 2013;1139–1147.

32. Saleh B, Elgammal A. Large-scale classification of fine-art paintings: learning the right metric on the right feature. arXiv preprint arXiv:1505.00855. 2015;1–21.

33. Tan WR, Chan CS, Aguirre HE, Tanaka K, Ceci n'est pas une pipe: a deep convolutional network for fine-art paintings classification. IEEE international conference on image processing (ICIP). IEEE. 2016;2016:3703–7.

34. Zhong S, Huang X, Xiao Z. Fine-art painting classification via two-channel dual path networks. Int J Mach Learn Cybern. 2020;11:137–52.

35. Sheng JC. Automatic categorization of traditional Chinese paintings based on wavelet transform. Comput Sci. 2014;41(2):317–9.

36. Li J, Yao L, Hendriks E, Wang JZ. Rhythmic brushstrokes distinguish van Gogh from his contemporaries: findings via automated brushstroke extraction. IEEE Trans Pattern Anal Mach Intell. 2011;34(6):1159–76.

37. Jiang W, Wang X, Ren J, Li S, Sun M, Wang Z, et al. MTFFNet: a multi-task feature fusion framework for Chinese painting classification. Cogn Comput. 2021;13:1287–96.

38. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen L C. Mobilenetv2: inverted residuals and linear bottlenecks. Proceedings of the IEEE conference on computer vision and pattern recognition. 2018;4510–4520.

39. Chollet F. Xception: Deep learning with depthwise separable convolutions. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017;1251–1258.

40. Dumoulin V, Visin F. A guide to convolution arithmetic for deep learning. arXiv preprint arXiv:1603.07285, 2016;1–31.

41. Lin M, Q Chen, S Yan. Network in network. arXiv preprint arXiv: 1312.4400 v3, 2013.

42. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. Proceedings of the IEEE conference on computer vision and pattern recognition. 2015:1–9.

43. Yan Y, Ren J, Sun G, Zhao H, Han J, Li X, et al. Unsupervised image saliency detection with Gestalt-laws guided optimization and visual attention based refinement. Pattern Recogn. 2018;79:65–78.

44. Chen R, Huang H, Yu Y, Ren J, Wang P, Zhao H, et al. Rapid detection of multi-QR codes based on multistage stepwise discrimination and a compressed MobileNet. IEEE internet of things journal. 2023;1–15.

45. Xie G, Ren J, Marshall S, Zhao H, Li R, Chen R. Self-attention enhanced deep residual network for spatial image steganalysis. Digital signal processing, 2023; Early Access.