

DSSESKN: A depthwise separable squeeze-and-excitation selective kernel network for art image classification

Shaojie Zhang^{1,*}

¹School of Art and Design, Zhengzhou University of Industrial Technology, Zhengzhou 450000 China

Abstract

Image classification is one of the key technologies of content-based image retrieval, and it is also the focus and hotspot of image content analysis research in recent years. Through the image processing and analysis technology to automatically analyze the image content to complete the management and retrieval of images, this process is the main content for image classification. Faced with massive digital Chinese art works, how to achieve effective management and retrieval for them has become an urgent problem to be solved. Traditional image retrieval technology is mainly based on image annotation, which has many problems, such as large workload and not objective enough. In this paper, we propose a depthwise separable squeeze-and-excitation selective kernel network (DSSESKN) for art image classification. SKNet (Slective Kernel Network) is used to adaptively adjust the receptive field to extract the global and detailed features of the image. We use SENet (squeeze-and-excitation network) to enhance the channel features. SKNet and SENet are fused to built the DSSESKN. The convolution kernel on the branch of DSSESKN module is used to extract the global feature and local detail features of the input image. The feature maps on the branches are fused, and the fused feature maps are compressed and activated. The processed feature weights are mapped to the feature maps of different branches and feature fusion is carried out. Art images are classified by deep separable convolution. Finally, we conduct experiments with other state-of-the-art classification methods, the results show that the effectiveness of the DSSESKN obtains the better effect.

Keywords: art image classification, depthwise separable, squeeze-and-excitation, selective kernel network, feature map.

Received on 19 November 2021, accepted on 24 November 2021, published on 26 November 2021

Copyright © 2021 Shaojie Zhang *et al.*, licensed to EAI. This is an open access article distributed under the terms of the [Creative Commons Attribution license](#), which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi: 10.4108/eai.26-11-2021.172304

*Corresponding author. Email: 352720214@qq.com

1. Introduction

Natural art painting is the embodiment of people's pursuit for life quality and spiritual quality. It can express emotions that cannot be expressed in words, including printmaking, Chinese painting, oil painting, gouache painting, watercolor painting, etc., as shown in figure 1.

Printmaking is a work of art produced by artists through conception creation, engraving and printing [1]. It expresses the theme style in the form of full and dense composition and small, simple composition. Chinese painting uses a brush dipped in water, ink or color on silk or paper to draw, pays attention to the spirit, the image of the object shape and bone method of the brush. Among them, the bone method expresses the artistry of using the

pen, including the pen force, the sense of force, the performance structure, etc. Chinese painting does not pay attention to the rendering of the background of the image [2]. Most of the paintings have blank space, and ink is used instead of color, so that the ink produces rich and subtle color changes, that is, "the ink is divided into five colors". The speed of the brush and the different length of the brush make the ink skills ever-changing. The light and dark tones are rich and colorful. Oil painting is different from Chinese painting in that it focuses on realism. The painter creates the sense of light through the contrast between warm and cold tones, the intensity of light and shade, and the thickness and thickness of pigments. Oil painting has opacity, rich color and high saturation [3]. Watercolor painting is a painting form in which transparent pigments are mixed with water as the medium. The transparency of pigments makes the picture transparent and clear, while the fluidity of water enhances the sense of fluidity, natural and free. Gouaches are a form of painting in which water is mixed with silty pigments. The color transparency is between watercolor and oil painting. Gouache is in wet when the color saturation is very high, dry color loses luster, the saturation is decreased. Oil painting, watercolor and gouache can show the characters with rich and colorful color, rough texture, obvious color block. However, the interaction of light and dark tones and the use of light and shadow in the oil painting make the overall sense of space and three-dimensional more intense. The details are rich in texture. The lines formed by the repeated stacking of pigments enhance the semantic information of the oil painting, and the transparency is the lowest. Gouache transparency followed, gouache dry, the picture appears obvious powder color block, three-dimensional sense is not strong, color saturation is lower than oil painting. The transparency of watercolor painting is high, which can express the fresh and lively texture. Water stains and wet marks on the picture are one of the key features to distinguish them. Chinese painting and watercolor painting have similarities in aesthetic expression, intention and charm, but on the whole, Chinese painting uses ink painting system, rich lines, details in the form of points, lines, surface to describe the object's shape, bone method, light and dark and modal verve, the overall color of the picture is deep. The overall color of watercolor painting is rich, bright and moist, and the brightness is large. The details do not pay attention to the line painting, and the interaction between "water" and "color" is used to express the mood and charm of the picture.



Figure 1. Five types of artistic images. (a) printmaking; (b) Chinese paintings; (c) oil paintings; (d) opaque watercolor paintings; (e) watercolor paintings

At present, the research work of art image management includes paintings classification according to subject matter and technique of expression, painters classification according to their creative styles, art portrait true or false identification. There are few studies on the classification of multi-category art images based on their style features. Different painters use different strokes, and the thickness of the lines in a painting can show the painter's style. Li [4] proposed to use 2-dimensional multi-resolution hidden Markov hybrid model to analyze most regions of the image, capture stroke features of key regions, and classify, compare painters. In traditional Chinese paintings, ink is often used instead of color. Sheng [5] used gray histogram to extract overall style features of ink paintings, used Sobel edge detection method to obtain local detail features with rich stroke styles, and used information entropy fusion algorithm to realize painter classification. Shen [6] divided the image into 4×4 sub-blocks. For each sub-block, it extracted the whole features such as color, texture, shape and the local texture features. They were input to radial basis function (RBF) neural network to train the parameters, then the vector of RBF neural network output hamming distance calculation in order to realize the painter classification of input image. Liang et al. [7] used monte Carlo convex hull feature selection model and support vector machine to describe the features of traditional Chinese paintings. According to the texture, color, shape and other features of Chinese painting, Wang [8] adopted the traditional method to extract the supervised heterogeneous sparse features, but the features only had 96 dimensions, which was insufficient to describe the overall features of Chinese painting. Jiang et al. [9] classified traditional Chinese fine brushwork and freehand brushwork based on the differences in expression techniques, and made use of the texture, shape, color and edge features at the lower level of images. Liang et al. [10] used scale invariant feature transform

(SIFT) feature detector and edge detection to get the key regions of the image. By describing the features of the key regions and the differences within the neighborhood, they analyzed the differences in expression techniques between fine brushwork and freehand brushwork via the cascading classifier. However, the characteristics of art are often combined in an organic form. It is difficult to generalize the combination skills. Different types of art images are similar in the features such as color, texture, line and so on. Extracting the texture, color, line and other features of art images using traditional methods cannot fully distinguish the style characteristics of each type of art image.

Convolutional neural networks (CNNs) [11-13] have achieved excellent results in extracting the global features and local details. Different convolution kernels in the branch of InceptionV4 module extract the overall and detailed features of the input image, but most of these modules consider the spatial dimension information without considering the dependence between channels, and it does not further enhance the extracted features. The SE (Squeeze-and-excitation) module in SENet (Squeeze-and-excitation networks) fuses the dependency between spatial dimensions and channels, strengthens the extracted useful features, and inhibits the irrelevant features. The SK (Selective kernel) module of SKNet enables the network to adaptively adjust the size of receptive field according to multiple scales of input information, and extracts the details and overall features of images. The SE and SK modules are shown in figure 2.

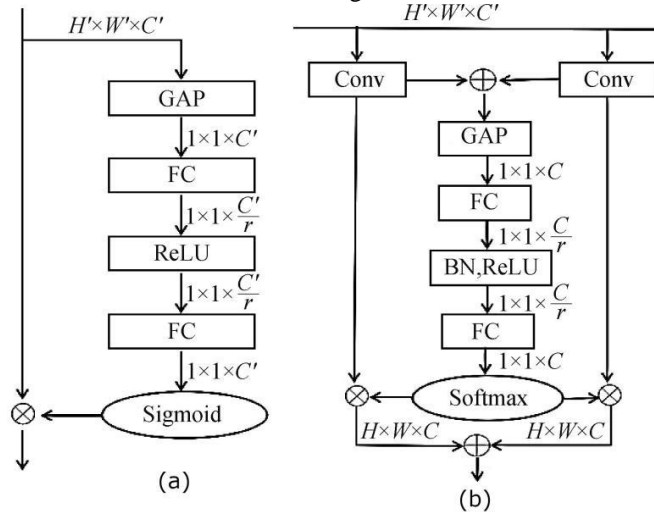


Figure 2. SE module and SK module. (a) SE. (b) SK.

This paper constructs DSSESKN module according to the characteristics of SE and SK modules. In the branch of DSSESKN module, the convolution kernel with different sizes extracts the overall and detailed features of the

image. And it executes fusion, compression, and activation. Then the extracted style features are weighted and mapped to the feature maps on each branch. Finally, the different feature maps are fused to strengthen the feature information extracted by different convolution kernels. In this paper, DSSESKN network is used to classify five types of art images.

2. Art image classification based on DSSESKN

2.1. DSSESKN module

In this paper, DSSESKN combines the SE module and the SK module to better enhance the overall style features and local detail features of the extracted art images. The sub-modules Split, Squeeze, Scale 4 and Scale are shown in figure 3. The expression is:

$$V = \sum_{i=1}^N U_i \cdot F_{ex}(F_{sq}(F_{gp} 9U)) \quad (1)$$

Where, U is the feature graph after fusion on the branch of DSSESKN module. $F_{gp}(\cdot)$ indicates the global average pooling (GAP) operation [14]. $F_{sq}(\cdot)$ represents channel compression processing. $F_{ex}(\cdot)$ is a channel feature activation operation. N is the number of branches in DSSESKN module. In this paper, $N=2$.

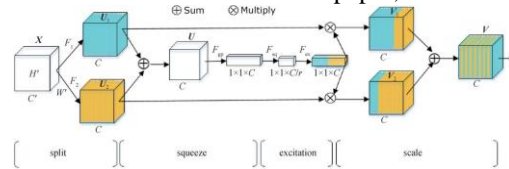


Figure 3. DSSESKN module

1) Split. For an intermediate feature graph $X \in R^{H' \times W' \times C'}$, two convolution kernels with different sizes are used for convolution mapping, i.e., $F_1 : X \rightarrow U_1 \in R^{H \times W \times C}$; $F_2 : X \rightarrow U_2 \in R^{H \times W \times C}$.

Where, H' , W' and C' represent the height, width and channel number of the feature graph X respectively. F_1 and F_2 are mapping processes through convolution kernel convolution mapping, batch normalization (BN) [15], and ReLU excitation function respectively. H , W and C represent the height, width and channel number of the feature map after F_1 and F_2 . Convolution filter $w = [w_1, w_2, \dots, w_c, \dots, w_C]$, where w_c represents the

parameter of the c -th filter. On the i -th branch, the convolution mapping formula of each filter to the intermediate feature graph X is:

$$u'_{ic} = w_c \times X = \sum_{k=1}^{c'} w_{ic}^k \times X^k + b_{ic} \quad (2)$$

$$u_{ic} = \delta(\zeta(u'_{ic})), i=1, \dots, N$$

Where, c' represents the number of channels in the filter and feature map, and b_{ic} represents bias. $\zeta(\cdot)$ represents batch normalization. $X' = \zeta(u'_{ic})$,

$$U_i = [u_{i1}, u_{i2}, \dots, u_{iC}].$$

2) Squeeze. After the split operation, two new feature graphs U_1 and U_2 are obtained. By summing the corresponding elements, the feature information of the two feature graphs is fused, i.e.,

$$U = U_1 + U_2 \quad (3)$$

The fused feature graph U combines the feature information of U_1 and U_2 , and uses global average pooling to pool the global spatial feature nodes in each feature channel. Global average pooling compresses feature graph U space information into C channel descriptors, and generates statistic $S \in R^C$ describing feature channel information. The c -th element in the statistic S is calculated by squeezing the space information of U , which is,

$$S_c = F_{gp}(U_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W U_c(i, j) \quad (4)$$

3) Excitation. In order to enhance the style features extracted from each type of art image and reduce the feature information with little effect, 1×1 convolution operation is used to reduce the channel dimension of the feature map after global average pooling to $1/r$ of the original channel, where r is the decline rate. After BN processing and ReLU function, the channel number is raised to the original number by 1×1 convolution operation. Finally, the normalized weight between (0,1) is obtained through a gate mechanism of Sigmoid, and the style feature information of each type of art image is obtained.

$$Z = F_{ex}(S) = \sigma(W_2 F_{sq}(S)) = \sigma(W_2 \delta(\zeta(W_1 S))) \quad (5)$$

Where $\sigma(X'') = (1 + e^{-X''})^{-1}$ is the Sigmoid activation

function. $X'' = W_2 \delta(\zeta(W_1 S))$, $W_1 \in R^{\frac{C}{r} \times C}$,

$W_2 \in R^{\frac{C}{r} \times C}$. r represents the dimension reduction ratio of the feature graph channel.

4) Scale. The weight Z obtained through squeeze and excitation is the image feature information. The weighted sum is obtained by:

$$V_c = u_{1c} \cdot Z_c + u_{2c} \cdot Z_c \quad (6)$$

Where $V_c = R^{H \times W}$ denotes the feature information of the c -th channel of the feature graph V , $V = [V_1, V_2, \dots, V_C]$.

2.2. Depthwise separable convolution in DSSESKN

Referring to the network structure of MobileNetV1 [16], depthwise separable convolution is used to build the new network in this paper. The first layer uses the dilated convolution to extract the features of the original art image. Compared with ordinary convolution, dilated convolution has a larger receptive field and can keep more internal data structure and original image information under the condition that the computation amount does not increase. Depthwise separable convolution consists of depthwise convolution and pointwise convolution. The operation formula of DSSESKN module using depthwise separable convolution is:

$$Y(X) = [Y_1 \cdot Y_2 \cdot Y_3](X) \quad (7)$$

Here, $Y_1 : R^{H' \times W' \times C'} \rightarrow R^{\frac{H'}{s} \times \frac{W'}{s} \times C'}$ is the depthwise

convolution operation. $Y_2 : R^{\frac{H'}{s} \times \frac{W'}{s} \times C'} \rightarrow R^{\frac{H'}{s} \times \frac{W'}{s} \times tC'}$ is the pointwise convolution operation.

$Y_3 : R^{\frac{H'}{s} \times \frac{W'}{s} \times tC'} \rightarrow R^{H \times W \times C}$ is the operation of DSSESKN module. s represents dimension reduction of feature graph dimension by depthwise convolution. t represents the processing of pointwise convolution for the channel number in the feature graph. After each depthwise convolution and pointwise convolution operation, BN and ReLU excitation functions are executed.

Table 1 shows the operation situation of taking 3×3 and 5×5 for ordinary convolution on DSSESKN branch. In order to reduce the over-fitting in the training, the L2 regularization method is used in the pointwise convolution operation. Then the dropout is utilized before GAP processing. In table 1, dilated_conv indicates the dilated convolution, d=2 indicates that the expansion rate is 2. depthwise_conv is the depthwise convolution, and the dep_separable_conv indicates the depth separable convolution. s1 indicates that the step size is 1, and so on.

Table 1. The parameter in DSSESKN module.

I D	Input size	Convolution type	Kernel size
1	299×299× 3	dilated_conv/s1	3×3,32,d=2
2	299×299× 32	depthwise_conv/s3	3×3
3	99×99×32	dep_separable_con v/s1	3×3,64
4	99×99×64	dep_separable_con v/s2	3×3,128
5	50×50×12 8	dep_separable_con v/s1	3×3,128
6	50×50×12 8	DSSESKN_clock/s 1	3×3,5×5,128,r =4
7	50×50×12 8	dep_separable_con v/s2	3×3,256
8	25×25×25 6	dep_separable_con v/s1	3×3,256
9	25×25×25 6	DSSESKN_clock/s 1	3×3,5×5,256,r =4
1 0	25×25×25 6	dep_separable_con v/s2	3×3,512
1 1	13×13×51 2	dep_separable_con v/s1	3×3,512
1 2	13×13×51 2	DSSESKN_clock/s 1	3×3,5×5,512,r =4
1 3	13×13×51 2	DSSESKN_clock/s 2	3×3,1024
1 4	7×7×1024	dep_separable_con v/s1	3×3,1024
1 5	7×7×1024	Global average pool	7×7
1 6	1×1×1024	Fully connected	1024×5
1 7	1×1×5	softmax	----

3. Experiment and analysis

3.1. Experiment environment

Adam is used as the optimizer in the experiment. The initial value of learning rate is 0.001. The training cycle is 120 times. In the training, the training set is rotated randomly from 00 to 200, and is randomly flipped and shifted by 0 to 10% in the horizontal and vertical directions respectively to enhance the generalization ability of the network. When the accuracy of network training does not improve after 3 times of training, the learning rate drops to 10%. In the experiment, 80% of the images is randomly selected as the training set and the remaining 20% as the verification set. Experimental hardware: two NVIDIA TESLAP100-12G GPUs, two Intel Xeon E5-2620 V4 CPUs, four DDR4 2 400 MHz16 GB memory. The deep learning framework uses Keras.

3.2. Data sets

The art images used in this paper are obtained from websites such as Arlib World Art Appreciation Database [17] and <https://www.studentart.com.cn/>, including 3400 prints, 3500 Chinese paintings, 3400 oil paintings, 3400 gouache paintings and 3400 watercolors. The style information of all art images is distributed evenly. The images with high resolution and rich style information are cropped into several images with 299×299pixels, and the images with rich style information are filtered manually to enhance the data. After data enhancement, there are 5200 prints, 5200 Chinese paintings, 5120 oil paintings, 5130 watercolors and 5128 gouaches. The network model in table 1 is used to conduct experiments on the original data and enhanced data. The results are shown in table 2. It can be seen that the classification accuracy of enhanced data is improved by 12.03%.

Table 2. Classification results before and after data augmentation/%

Data augmentation	Accuracy
Without data augmentation	77.51
With data augmentation	89.54

3.3. Experiment comparison results

Results with different classification methods

1) Comparison with traditional network models.

In order to verify the effect of the DSSESKN in this paper on the classification of art images, the data sets are

input into the traditional network model and the DSSESKN model respectively for training and verification. The results are shown in table 3.

Table 3. Classification comparison with different models

Model	Parameter /M	Time/min (training+testing)	Accuracy /%
ResNet50	23.5	6041	85.86
Xception	20.8	4321	85.27
InceptionV4	41.7	6551	85.88
MobileNetV1	3.2	1656	86.01
MobileNetV2	2.2	1370	77.88
ShuffleNet	9.2	558	80.41
Proposed+SE(r=16)	2.2	1356	85.35
Proposed+SK(r=16)	8.4	2369	84.96
DSSESKN(r=16)	13.8	2139	86.21
DSSESKN(r=4)	14	1660	89.54

As can be seen from table 3, compared with the InceptionV4, ResNet50 and Xception network models, the DSSESKN model has light parameters, short training time and has high accuracy. However, the DSSESKN module contains parallel convolution operation, which makes the network model parameters in this paper higher than ShuffleNet, MobileNetV1 and MobileNetV2. Compared with Proposed+SE and Proposed+SK(r=16), the accuracy of DSSESKN is 86.21%, which improves by 0.86% and 1.25% than that of SE and SK, respectively. When r=4, the classification accuracy of DSSESKN is 89.54%, which is higher than that when r=16.

2) Comparison with traditional methods.

Traditional global feature and local feature extraction methods are used to classify the data in this paper, and the results are shown in table 4.

Table 4. Comparison with traditional methods

Methods	Accuracy/%
Reference[5]	66.79
Reference[6]	61.3
DSSESKN(r=4)	89.54

As can be seen from table 4, the traditional manual features and local features cannot fully distinguish the style features of various art images. However, the DSSESKN can better extract the overall features and local detail features of art images and improve the classification accuracy.

3) Drop rate and branch convolution kernel size.

Drop rate r and branch convolution kernel size are a set of important parameters to control computational resources and experimental accuracy in DSSESKN module. The r value of DSSESKN and the size of branch convolution kernel are tested and analyzed. The results are shown in table 5. It can be seen that when the size of the convolution kernel on the branch of the DSSESKN module is fixed, the classification result with r=4 is higher than that r=16. When r value is fixed, DSSESKN module with branch=2 has shorter training time and fewer parameters than that of branch=3. When the convolution kernel on the branch of DSSESKN module is 1×1 and 5×5 respectively, the experiment results have the highest accuracy.

Table 5. Comparison between drop rate and convolution kernel size

Drop rate	1×1	1×1	5×5	Parameter r/M	Time/min	Accuracy/%
r=4	√	√		2.3	1368	87.28
	√		√	2.6	1681	87.59
		√	√	2.7	1811	87.24
	√	√	√	2.7	2370	87.36
r=16	√	√		2.3	1230	86.36
	√		√	2.6	1220	86.87
		√	√	2.7	1440	86.15
	√	√	√	2.7	1770	86.42

4) Dilated convolution.

In order to compare the effect of dilated convolution on the feature extraction of art images, the data with different sizes of dilated convolution are made experiments, and

the results are shown in table 6. Where K1 denotes the ordinary 3×3 convolution kernel. K2 represents 3×3 convolution kernel with expansion rate=2 and receptive field= 5×5 . K3 represents 3×3 convolution kernel with expansion rate=3 and receptive field= 7×7 .

Table 6. Classification results with different dilated convolution kernels

Dro P rate	K	K	K	Parameter /M	Time/m in	Accuracy /%
	1	2	3			
r=4	√	√		2.4	1530	86.08
	√		√	2.4	1657	86.01
		√	√	2.4	1560	84.51
	√	√	√	2.6	2220	86.54

The experiment results show that the parameters of dilated convolution are less than that of ordinary convolution with the same receptive field, but the classification accuracy is not as high as that of ordinary convolution.

Feature visualization

In order to facilitate the observation of key region features of network model, we select proposed+SE, proposed+SK and DSSESKN to make comparison. The Grad-CAM (gradient-weighted class activation) is used to visually display the extracted feature areas of the art images as shown in figure 4.

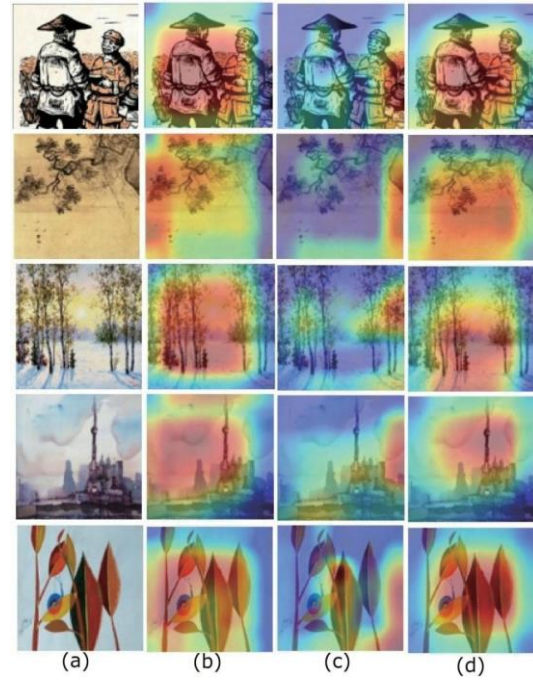


Figure 4. Characteristic distribution thermal diagram. (a) original paintings; (b) DSSESKN; (c) proposed+SK; (d) proposed+SE.

Figure 4(a) is an art image. Figure 4(b)-(d) is feature thermal map. The deeper warm color denotes that the category judgment depends on the information in this region greatly. As can be seen from figure 4, the style features of art images are relatively uniform in the overall distribution. Compared with the proposed+SE and proposed+SK, the DSSESKN extracts the overall features of art images more comprehensively, and the local details are more prominent.

Analysis of classification prediction results



Figure 5. Classification results of five types of art images. (a) printmaking; (b) Chinese painting; (c) oil painting; (d) opaque watercolor painting; (e) watercolor painting.

To visually demonstrate the classification performance of the proposed method, the images in figure 5 are classified using the DSSESKN network. In figure 5, the blue box indicates that the oil painting is wrongly classified as gouache painting, the green box indicates that the oil painting is wrongly classified as watercolor painting, and the red box indicates that the gouache is wrongly classified as oil painting. As can be seen from figure 5, oil painting, gouache painting and watercolor painting have certain similarities in painting techniques, color layout, texture, application of warm and cold tones or design of light sensitivity. Oil painting and gouache painting that are classified as watercolor painting and gouache painting are not obvious in the style characteristics of their categories.

Classification performance comparison

The recall, precision and F-measure values are used to evaluate the performance of DSSESKN, and the results are shown in table 7. The accuracy and F-measure values of traditional Chinese painting are the highest, and the recall of printmaking is the highest. Printmaking’s unique

carving techniques and line characteristics are the key features that distinguish them from other kinds of paintings. Chinese paintings’ unique painting skills and the use of ink color make it different from western paintings.

Table 7. Classification performance/%

Art images	Recall	Precision	F-measure
Printmaking	87.36	91.14	88.94
Chinese painting	93.19	87.43	90.22
oil painting	88.26	84.55	86.37
opaque watercolor painting	87.33	86.55	86.94
watercolor painting	82.12	87.89	84.46

Receiver operating characteristic (ROC) curve and area under curve (AUC) area are the main indicators to evaluate the performance of a classifier. The larger ROC curve denotes the better classification performance. The classification performance of DSSESKN is measured using ROC and AUC, as shown in figure 6. It can be seen that the AUC value of Printmaking reaches 0.99, and the classification result of classifiers for Printmaking is the best, followed by Chinese paintings. Because the color blocks, lines, painting techniques and other features of oil painting are similar to other paintings in overall features and local details, the AUC value of oil painting is the smallest.

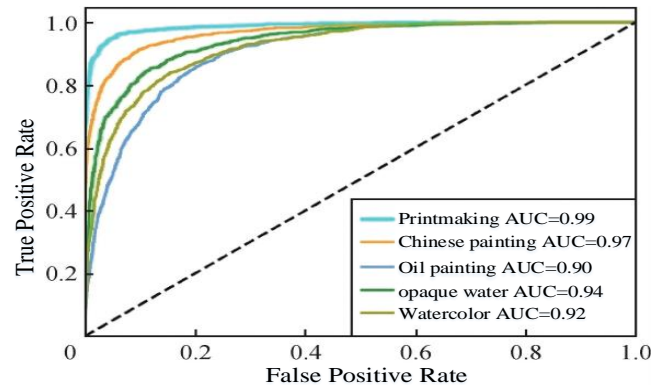


Figure 6. ROC curve

4. Conclusion

This paper proposes a new DSSESKN model for art image classification. The DSSESKN module uses the two-branch convolution kernel to extract the overall feature and local detail features of the image, and fuses the extracted features. Two 1×1 convolution operations and Sigmoid gate mechanism are used to extract the key features of each art image. The depthwise separable convolutional neural network is constructed to classify art images such as printmaking, Chinese painting, oil painting, watercolor painting and gouache painting, which can better realize the feature extraction and classification of art images. The DSSESKN achieves better classification effect of art images than existing network models. However, the parallel double convolution operation in the DSSESKN module increases the number of network model parameters significantly. In the following work, we will optimize the network model of art image classification, enhance the art images datasets, and further improve the classification accuracy of art images.

Acknowledgements.

The authors thank reviewers for their anonymous comments.

References

- [1] X. Yang and H. Zhang. Art image classification with double kernel squeeze-and-excitation neural network[J]. *Journal of Image and Graphics*, 25 (05) :0967-0976, 2020.
- [2] Zhao S, David K, Simon B, et al. Enhancing the Appreciation of Traditional Chinese Painting Using Interactive Technology[J]. *Multimodal Technologies & Interaction*, 2018, 2(2):16-.
- [3] Zhao S. An Analysis of Interactive Technology's Effect on the Appreciation of Traditional Chinese Painting: A Review of Case Studies[J]. *The International Journal of New Media, Technology and the Arts*, 14(3):1-12, 2019.
- [4] Jia Li and J. Z. Wang, "Studying digital imagery of ancient paintings by mixtures of stochastic models," in *IEEE Transactions on Image Processing*, vol. 13, no. 3, pp. 340-353, March 2004, doi: 10.1109/TIP.2003.821349.
- [5] Sheng J, Jiang J. Style-based classification of Chinese ink and wash paintings[J]. *Optical Engineering*, 2013, 52(9):093101.
- [6] Shen J. Stochastic modeling western paintings for effective classification[J]. *Pattern Recognition*, 2009, 42(2):293-301.
- [7] Ye Liang, Hong Bao and Hong-Zhe Liu, "A simple method for classification of traditional Chinese Painting and Calligraphy images," 2010 International Conference on Educational and Information Technology, 2010, pp. V3-340-V3-344, doi: 10.1109/ICEIT.2010.5608356.
- [8] Wang Z, Sun M, Han Y, et al. Supervised Heterogeneous Sparse Feature Selection for Chinese Paintings Classification[J]. *Journal of Computer-Aided Design and Computer Graphics*, 2013, 25(12):1848-1855.
- [9] Jiang S Q, Huang Q M. Ye QX and Gao W. An effective method to detect and categorize digitized traditional Chinese paintings. *Pattern Recognition Letters*, 27(7): 734-746, 2006.
- [10] Liang S, Huang Y, Li S, et al. Automatic traditional Chinese painting classification: A benchmarking analysis[J]. *Computational Intelligence*, 2020, 36(3), Pages 1183-1199.
- [11] T. Gao, H. Li, and S. Yin. Adaptive Convolutional Neural Network-based Information Fusion for Facial Expression Recognition [J]. *International Journal of Electronics and Information Engineering*. Vol. 13, No. 1, pp. 17-23, 2021.
- [12] Yin, S., Li, H. & Teng, L. Airport Detection Based on Improved Faster RCNN in Large Scale Remote Sensing Images [J]. *Sensing and Imaging*, vol. 21, 2020. <https://doi.org/10.1007/s11220-020-00314-2>
- [13] X. Wang, S. Yin, K. Sun, et al., GKFC-CNN: Modified Gaussian Kernel Fuzzy C-means and Convolutional Neural Network for Apple Segmentation and Recognition [J]. *Journal of Applied Science and Engineering*, vol. 23, no. 3, pp. 555-561, 2020.
- [14] Yang S, Nian F, Wang Y, et al. Real-time face attributes recognition via HPGC: horizontal pyramid global convolution[J]. *Journal of Real-Time Image Processing*, 2020, 17(10).
- [15] X. Wang, S. Yin, H. Li. A Network Intrusion Detection Method Based on Deep Multi-scale Convolutional Neural Network[J]. *International Journal of Wireless Information Networks*. 27(4), 503-517, 2020.
- [16] Zhou B, Jin S, Liu Y, et al. OpenCL-code generation framework for MobileNets with depthwise separable convolution and merged layers[C]// *Parallel Processing of Images and Optimization Techniques*. 2020.
- [17] E. J. Umlauf, H. Piringer, G. Reitmayr and D. Schmalstieg, "ARLib: the augmented library," *The First IEEE International Workshop Augmented Reality Toolkit*, 2002, pp. 2 pp.-, doi: 10.1109/ART.2002.1106996.