

Research Article

Sentiment Analysis of Chinese Paintings Based on Lightweight Convolutional Neural Network

Jiaying Bian  and Xiaoying Shen 

Wuxi Vocational College of Science and Technology, No. 8 Xinxu Road, Wuxi, Jiangsu 214000, China

Correspondence should be addressed to Xiaoying Shen; 1201801@wxsc.edu.cn

Received 13 July 2021; Revised 28 July 2021; Accepted 30 July 2021; Published 10 August 2021

Academic Editor: Shan Zhong

Copyright © 2021 Jiaying Bian and Xiaoying Shen. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Chinese painting is one of the representatives of our country's outstanding traditional culture, and it embodies the long history and intellectual wisdom of the Chinese nation. In the paper, we combine the artistic characteristics of Chinese paintings and use an optimized SqueezeNet model to study the sentiment analysis of Chinese paintings. To make full use of the advantages of lightweight convolutional neural networks, we make two optimizations based on SqueezeNet. On the one hand, expand the model width to obtain more effective Chinese painting sentiment features for classification tasks, thereby improving the classification accuracy of the model. On the other hand, introduce the idea of residual network to prevent gradient disappearance and gradient explosion in the training process, thereby enhancing the model's generalization ability. To verify the effectiveness of the optimized SqueezeNet model used in the sentiment analysis of Chinese paintings, four kinds of sentiment classifications were carried out on the multitheme Chinese paintings downloaded on the Internet. The results of comparative experiments show that the optimized SqueezeNet model used in this paper can improve the accuracy of classification and has better generalization ability. Finally, the research results of this paper can be applied to the protection of traditional culture, the appreciation of traditional Chinese painting, and art education and training, which is conducive to the inheritance and innovation of the national quintessence and promotes the prosperity and development of traditional art and culture.

1. Introduction

Traditional Chinese painting is the artistic treasure of our country. It not only vividly depicts the long history of our country with superb artistic skills but also embodies the ideology, philosophy, cultural concept, and aesthetic characteristics of the Chinese nation. It can be said that Chinese painting is an irreplaceable and important part of the traditional cultural and spiritual heritage of the Chinese nation and stands in the gallery of paintings in the world with its unique and distinctive artistic style. With the rapid development of technology, a large number of digital Chinese paintings have appeared in the Internet and digital museums. However, how to efficiently use and manage these paintings and then to promote the charm of Chinese paintings has become an urgent problem to be solved. Studying the algorithms based on sentiment analysis of Chinese paintings will not only help improve users' ability to learn and appreciate

Chinese paintings but also help promote the construction of digital museums and cultural relic management. It also helps to demonstrate cultural self-confidence and promote the implementation of the cultural power strategy, thereby comprehensively enhancing the soft power and international influence of our country's culture.

Different from Western painting's "write shape by shape," traditional Chinese painting emphasizes imagery. The author often creatively interprets the object artistically in the paintings. It can be said that Chinese painting has been an art form centered on emotion since ancient times. Excellent Chinese painting works are the spiritual product of the artist's depiction with true feelings. The painter does not simply describe the objective things but more importantly expresses his inner feelings, so as to achieve a state of combining things with himself and sublimating the spirit. Because of the artistic characteristics of traditional Chinese paintings, which emphasize "freehand brushwork" and

“focusing on emotional expression,” we inevitably need to consider the emotional factors in the painting when we appreciate, research, and protect Chinese paintings. Understanding high-level emotional semantics can help us better appreciate and study Chinese paintings. Meanwhile, it can also enrich the algorithms on organization, management, and retrieval of Chinese paintings and popularize the dissemination of Chinese paintings.

In view of this, in order to fully understand the emotions contained in Chinese paintings, scholars have done a lot of research work and achieved fruitful research results. Literatures [1, 2] extract local and global features based on histograms to characterize different aspects of the artistic style of Chinese paintings and use these features to drive neural networks to classify ink paintings. Literature [3] proposed an art descriptor based on the characteristics of composition and painting objects, dealing with the correlation and synergy among all the elements in the integrated features, and then combined the Monte Carlo convex hull feature selection model to classify the authors of Chinese paintings. Literature [4] used the STASM algorithm to extract the feature points of the face and the face in Chinese paintings and then converted the style of the user’s facial photo and integrated it into the Chinese painting template to form the user’s ink portrait. Literature [5] used a deep learning model to classify the authors of traditional Chinese ink paintings. They took the stroke image of each input painting as the recognition basis and then input all the strokes into the CNN-based feature extractor to form the combined high-dimensional features of each Chinese painting. Literature [6] used a hyperspectral camera with a specific frequency of visible light to scan Chinese paintings and combined the principal component analysis algorithm and the spectral and spatial features extracted by CNN to identify the authenticity of Chinese paintings. Literature [7] introduced the bottleneck layer idea of the Inception module in GoogLeNet on the basis of the deep convolutional encoder-decoder, in order to achieve the purpose of reducing model parameters and speeding up the calculation. Literature [8] proposed an algorithm to simulate the creative process of color ink painting. This algorithm uses CNN and Generative Adversarial Networks (GAN) to convert line and color styles and stylizes flower photos into color ink paintings. Literature [9] used Sobel edge detection to obtain the information of Chinese brush painting and then transformed the information by discrete cosine transform as the input of CNN and finally combined support vector machine to classify the two styles of Chinese painting meticulous and freehand. This algorithm establishes a hybrid model composed of CNN and support vector machine and only uses the stroke features in the painting. Literature [10] proposed an algorithm of joint mutual information and data-embedded classification. They first used the VGG model to extract the features of Chinese paintings and then used mutual information theory to make the distribution information of the image affected by the importance of the features, thereby improving the classification accuracy. Literature [11] proposed a style transfer algorithm for Chinese paintings. It mainly inputs the four restriction conditions of ink painting strokes, white space, ink smearing, and yellow-

ish tone into CNN, designs corresponding conversion strategies for different styles of freehand brushwork and meticulous brushwork, and obtains better ink painting visual effects.

The above research shows that the deep network structure can obtain more abstract high-level semantic features; it can handle more complex classification tasks. Therefore, deep convolutional network has obvious advantages in tasks such as feature extraction and classification. In order to improve the classification efficiency of the CNN in the application of Chinese painting sentiment classification, we use an optimized lightweight CNN to recognize the sentiment of Chinese paintings. The related research work is as follows:

- (1) Firstly, the theory of image sentiment analysis and related algorithms are studied to provide a theoretical basis for follow-up research
- (2) Secondly, combining the artistic characteristics of Chinese paintings, try to use a lightweight CNN to study the sentiment classification of Chinese paintings. At the same time, the researchers also made two optimizations based on SqueezeNet. On the one hand, expand the model width to obtain more effective Chinese painting sentiment features for classification tasks, thereby improving the classification accuracy of the model. On the other hand, introduce the idea of residual network to prevent gradient disappearance and gradient explosion in the training process, thereby enhancing the model’s generalization ability
- (3) Finally, the optimized lightweight CNN is applied to the sentiment recognition of Chinese paintings. The experimental results show that compared with other deep learning algorithms, the classification accuracy and efficiency of the algorithm used in this paper are relatively high in the task of Chinese painting sentiment classification, verifying the feasibility and effectiveness of the algorithm

2. Related Work

2.1. Image Sentiment Analysis. At present, most sentiment analysis tasks are based on text content, and many scholars have conducted comprehensive and in-depth research in the field of text sentiment analysis. However, as a special image classification technology, image sentiment analysis is still under constant exploration in both technical and application aspects. Different from text sentiment analysis, the emotion in the picture is often hidden, which is a kind of high-level semantic emotional understanding. Studying the sentiment of massive images can not only understand the emotional needs of users but also help optimize user experience in image retrieval and recommendation systems. What is more, it is also helpful for studying hot social issues, learning about the attitudes of the public, and providing data support for online public opinion analysis and monitoring.

Image sentiment analysis is a multidisciplinary task. Since the understanding and classification of emotions in pictures are affected by subjective, objective, and cultural

factors in aesthetics, image sentiment analysis involves many disciplines such as artificial intelligence, computer vision, psychology, and aesthetics. Image sentiment analysis mainly refers to the fact that the computer recognizes the content expressed in digital images to cause people's emotional responses, and then, according to different emotional responses, the images can be classified to different classes. Generally, there is a big difference between the feature information extracted by the computer from the low-level visual data and the human emotion interpretation [12, 13]. We call the difference the emotional gap [14, 15]. We hope that the computer can understand the image in depth and accurately recognize the emotional semantics contained in the image. At the same time, on the basis of understanding the emotional semantics of the image, the sentiment classification of the image is carried out according to the different understanding of the emotional semantics of the digital image. In other words, image sentiment analysis is a high-level semantic understanding of the information conveyed by the image. It can bridge the huge emotional gap between low-level visual features and high-level emotions. Today, when the amount of information is increasing rapidly, image sentiment classification is helpful for image labeling and retrieval, which has great social and commercial value and has attracted widespread attention [16].

2.2. The Related Methods Based on Image Sentiment Analysis.

Image sentiment analysis methods mainly include methods based on traditional machine learning [17, 18] and methods based on deep learning [19, 20]. Traditional image sentiment analysis methods first need to extract different dimensional visual features from digital images and use machine learning algorithms and models to train the emotions implicit in the digital images through the mapping of the sentiment space model. For example, literature [21] constructed a color matching emotional word data set based on the relationship between color and emotion and then found the corresponding emotional word in the data set to predict the sentiment of the painting by creating a color spectrum for Western paintings. Literature [22] extracted facial expression action unit features and found the relationship between them to realize video sentiment recognition. Literature [23] used directed acyclic graph support vector machine for facial sentiment classification of connected features. The algorithm extracts facial expressions based on geometric and texture features and proves that simple feature stitching can significantly improve the efficiency of facial sentiment classification. However, traditional image sentiment analysis methods do not take into account the gap between the low-level visual features and the high-level emotional semantics. The extraction of visual features is an important prerequisite for effective sentiment analysis. There is no clear limit on the degree of influence of visual features of different dimensions on the emotion of an image.

With the continuous expansion of the application field of deep learning, researchers have tried to use deep learning algorithms to automatically extract image features and realize sentiment classification, which have achieved better sentiment prediction results. Literature [24] proposed the progressive CNN deep model on the basis of CNN and

constructed a large artificially labeled Twitter visual sentiment data set. Literature [25] integrated the crossresidual neural network into multitask deep learning to solve the classification problem of image objects and their sentiment. Literature [26] analyzed the performance of the CNN layer by layer and fine-tuned the CNN to be applied to image sentiment prediction tasks. It proves the effectiveness of deep network learning to recognize sentiment-related features of natural images. Literature [27] used binary classification to assist in multiclassification tasks of natural image sentiment. However, the data set requires two sets of sentiment labels, binary classification and multiple classification, which increases the workload. Literature [28] combined the visual attention mechanism guided by the saliency map with the CNN architecture to achieve better sentiment classification performance based on natural images. In order to maximize the extraction of features that can represent image emotions, literature [29] proposed a cropping method that uses a fully convolutional network to select emotional regions from an image and uses the interdependence of tags to construct a structured learning model. Aiming at the local image sentiment classification, literature [30] used the feature pyramid network to extract multilayer depth features to remove redundant nonemotional areas. Literature [31] proposed an 11-layer CNN model with visual attention to solve the problem of facial expression recognition. The network model extracts CNN features from face images, calculates the region of interest, and finally, uses the features in the resulting region to determine facial expressions. Literature [32] combined the color histogram and the bottom layer features of local binary pattern texture features with deep sentiment features to identify image sentiment. However, the multilevel feature extraction and fusion steps are scattered and unsystematic. Literature [33] combined the features extracted according to the art theory with CNN features and used support vector machines to recognize image sentiment. However, this algorithm requires manual annotations such as eye movement trajectories, which reduces the practical applicability.

To sum up, the limitation of image sentiment analysis is that the elements that affect sentiment judgment include not only the information of the objects in the image, but the scene of the image can also trigger different emotions. Moreover, different objects in the same image may represent different sentiment classifications, such as the semantic segmentation problem in image sentiment analysis, which needs to be further studied by researchers.

3. Sentiment Analysis Model of Chinese Painting Based on Lightweight Convolutional Neural Network

3.1. *Lightweight Convolutional Neural Network.* The lightweight network SqueezeNet was proposed by Stanford and Berkeley researchers in 2017 [34]. The lightweight CNN model redesigns the network structure on the basis of the existing CNN structure to achieve the goal of reducing the amount of parameters and reducing the computational complexity. The core building module of SqueezeNet is Fire

module, which mainly consists of two parts: squeeze layer and expand layer. The specific operations of the two layers are shown in Figure 1. The squeeze layer uses a 1×1 convolution kernel to convolve the input features. The main purpose of this is to reduce the number of channels of input features, that is, dimensionality reduction. The expand layer uses 1×1 convolution operation and 3×3 convolution operation, respectively, and then concatenates the convolution results. The combination of these two layers can effectively reduce the amount of parameters.

As shown in Figure 1, the input feature size of the Fire module is $H \times W \times M$, and the output feature size is $H \times W \times (e_1 + e_3)$. It can be seen that the spatial size of the feature data before and after is unchanged; only the number of channels, that is, the number of dimensions, is changed. First, suppose that a feature map with a size of $H \times W \times M$ passes through the squeeze layer; then, S_1 feature maps are obtained. That is, the space size of the feature data remains unchanged, the number of channels changes from M to S_1 , where S_1 is smaller than M , so as to achieve the compression effect. Here is a brief introduction to feature map. In each convolutional layer in the CNN, the data exists in three-dimensional form. We can think of it as a stack of many two-dimensional images, each of which is called a feature map. Then, input the feature data of size $H \times W \times S_1$ into the expand layer, and go through a 1×1 convolution operation and a 3×3 convolution operation, respectively. Then, the results of the two convolutions are spliced as the output of the Fire module; that is, an output feature of size $H \times W \times (e_1 + e_3)$ is obtained. The Fire module has three adjustable parameters: S_1 , e_1 , and e_3 . They, respectively, represent the number of convolution kernels and also represent the dimension of the corresponding output feature map. In SqueezeNet, $e_1 = e_3 = 4S_1$.

The core idea of SqueezeNet is to connect multiple Fire modules in a cascaded form. It can give full play to the characteristics of the Fire module and reduce the amount of parameters in the network. The structure of SqueezeNet is shown in Figure 2. SqueezeNet starts with an independent convolutional layer (conv1) and then cascades 8 Fire modules, namely, fire2-fire9. Meanwhile, end the cascade with a convolutional layer (conv10). Finally, the global average pooling layer is used instead of the fully connected layer for output. From the beginning to the end of the network, gradually increase the number of convolution kernels in each Fire module. And after the conv1, fire4, and fire8 layers, the maximum pooling with a step size of 2 is used, respectively, and the global average pooling is performed after conv10. That is, the pooling layer is placed in the back position. The purpose of this is to provide a larger activation map for the convolutional layer. Because the activation map retains a lot of information, the larger the activation map of the convolutional layer, the higher the classification accuracy of the network. It should be noted that the Relu activation function is used in both the squeeze layer and the expand layer of SqueezeNet. After the Fire9 module, the Dropout technology was used, and the output specifications of the Dropout depend on the dimensions of the corresponding convolutional layer.

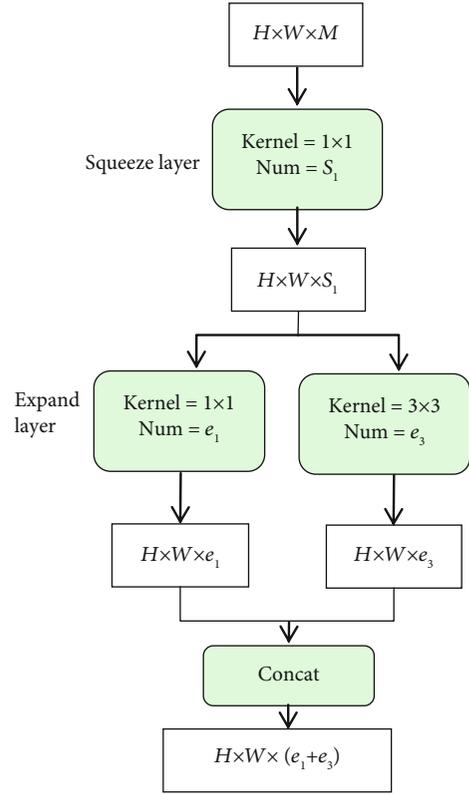


FIGURE 1: Specific operations of squeeze layer and expand layer.

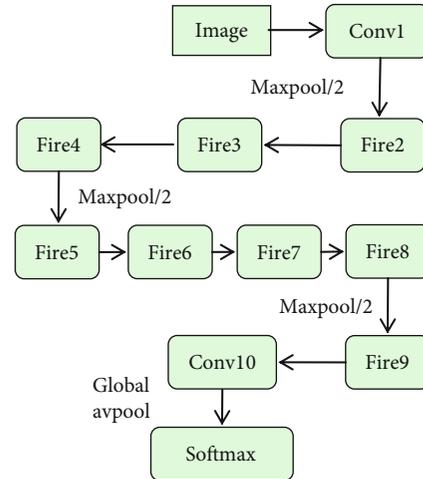


FIGURE 2: The structure of SqueezeNet.

In addition, in order to compress the network SqueezeNet, we also discarded the fully connected layer. Finally, the Softmax classifier is used to output the classification results, and the output specification depends on the classification classes. In the paper, we divide Chinese paintings into four emotions: sad and arrogant, arrogant, lively and cheerful, and quiet and peaceful. That is, the number of classes is 4.

3.2. ResNet. The proposal of residual network (ResNet) effectively solves the problem of gradient explosion or

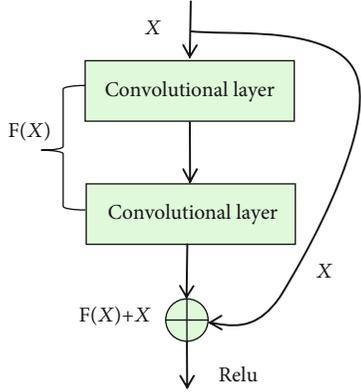


FIGURE 3: Residual error structure.

disappearance in deep network training. Although the deepening of the network structure can improve the accuracy of the network's classification of images, it is not feasible to blindly increase the depth of the network to improve the classification accuracy of the network. On the one hand, as the number of network layers deepens to a certain extent, its performance will not increase linearly with its depth. On the other hand, the deep network will have the problem of gradient disappearance or explosion during the training process. Therefore, deep networks are difficult to train, and ResNet was born to solve this problem. The core idea in ResNet is the residual idea, and the residual structure is shown in Figure 3.

The core idea of the residual structure is to introduce an "identical shortcut connection." For a convolutional layer, the new data $H(X)$ is obtained after the convolution operation of input data X . Now, add an identity mapping so that the function $H(X)$ is transformed into $F(X) + X$, where $F(X)$ is the residual. As the number of network layers increases, if a certain deep network model reaches the optimal level and continues to deepen the number of layers in the network model, even if the residual $F(X)$ approaches 0, the additional layers will only be equivalent. For the identity mapping operation, at least it will not affect the expressive ability of the model. Moreover, the residual error often does not become 0 during the actual training process, so the convolutional layer can further extract new and more abstract feature information, thereby further improving the representation ability of the model. The proposal of this idea makes it possible to increase the accuracy of the model by increasing the depth of the model.

3.3. SqueezeNet Optimization Design Based on Sentiment Classification of Traditional Chinese Paintings. In this section, we mainly optimize the lightweight network SqueezeNet in two parts. First, expand the width of the model. That is, the Fire module is added on the basis of the original network structure to improve the accuracy of model classification. Second, introduce the idea of residual network to prevent gradient disappearance and gradient explosion in the training process, thereby enhancing the generalization ability of the model for image classification and improving the efficiency of the model. The structure of the optimized

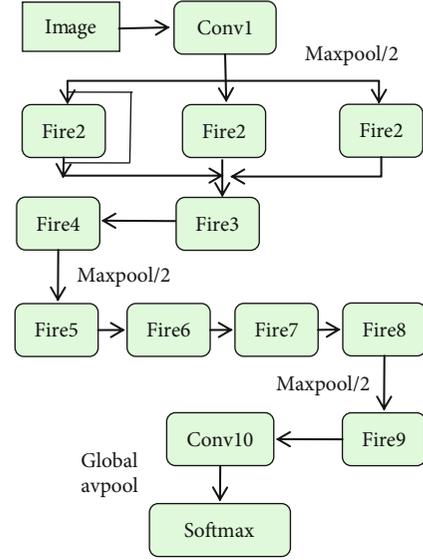


FIGURE 4: The structure of the optimized SqueezeNet based on sentiment classification of traditional Chinese paintings.

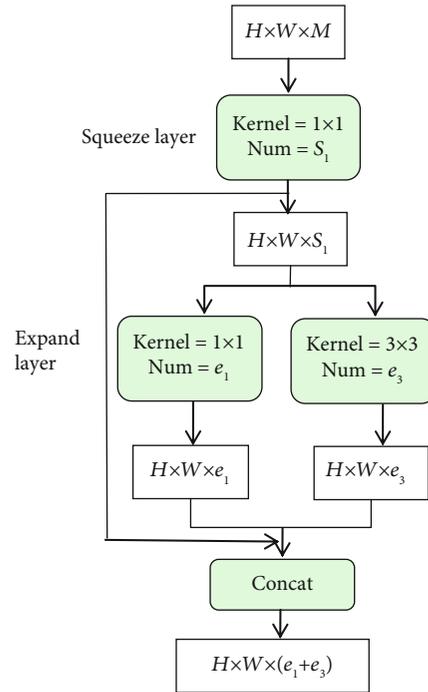


FIGURE 5: The improved Fire module structure.

SqueezeNet based on sentiment classification of traditional Chinese paintings is shown in Figure 4.

As shown in Figure 4, the optimized SqueezeNet did not increase the depth of the network but appropriately increased the width of the network. In addition, the idea of residual network is introduced into the Fire module, and the layer-jumping structure is added to prevent the gradient from disappearing during the training process, which is intended to improve the efficiency of the network. The improved Fire module structure is shown in Figure 5.

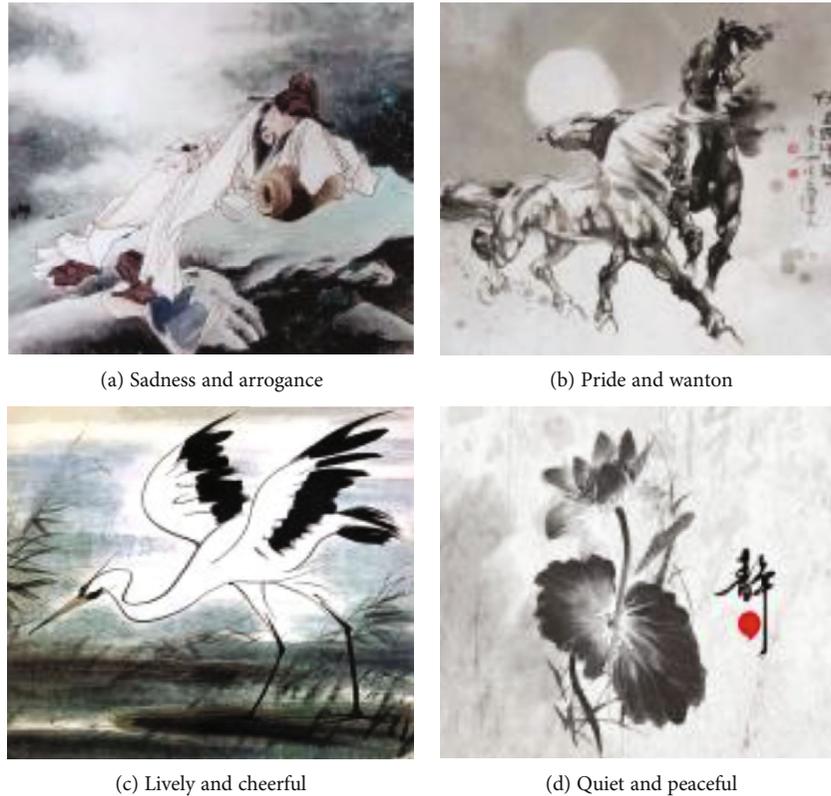


FIGURE 6: The representative works of different emotions.

4. Experimental Results and Analysis

4.1. Data Set. In this paper, we obtained Chinese paintings with multiple themes and periods from the Internet to form a Chinese painting sentiment data set. The data set collected a total of 800 samples of Chinese paintings, including portrait paintings, landscape paintings, flower and bird paintings, and animal paintings. The fine brushwork and freehand painting techniques are all involved. The sample data is extensive, and the quality of images is clear. Based on the analysis of the inscriptions and poems of the paintings, the painter's life experience, and the appreciation and comments of the works, the sample data is divided into four types of common emotions in traditional Chinese paintings: sadness and arrogance, pride and wanton, lively and cheerful, and quiet and peaceful. And the amount of data in each class is equal to ensure the balance of the data set. Figure 6 shows the representative works of different emotions.

4.2. Experimental Results and Analysis. In the experiments, we use classification accuracy, network parameters, and network calculations as evaluation indicators to verify the performance of the optimized SqueezeNet in the sentiment analysis of traditional Chinese paintings. Meanwhile, the sentiment class marked with the guidance of the expert is regarded as the correct label. The amount of network calculations refers to the number of floating-point operations required to infer an image. The network calculation amount is obtained by the corresponding relationship among the number of input channels, the number of output channels,

TABLE 1: Classification accuracy of different network models.

Network models	Classification accuracy
AlexNet	0.76
ResNet	0.82
SqueezeNet	0.74
Optimized SqueezeNet	0.86

the height and width of the convolution kernel, and the height and width of the output channels.

First of all, the optimized SqueezeNet and AlexNet, ResNet, and SqueezeNet are, respectively, used to perform classification accuracy experiments on the test set. The classification accuracy of different network models is shown in Table 1.

Analyzing the experimental results in Table 1, it is found that the classification accuracy of the other three models is lower than that of the optimized SqueezeNet on the task of Chinese painting sentiment recognition, since the network model used in the paper is optimized on the basis of SqueezeNet. By expanding the width of the network model and slightly increasing the model parameters, we can obtain more effective Chinese painting sentiment features for classification tasks, thereby improving the classification accuracy of the model. Therefore, the optimized SqueezeNet has a better classification effect and is more suitable for sentiment analysis of Chinese paintings.

Finally, the network parameters and calculations of AlexNet [35], ResNet [36], SqueezeNet [34], and the

TABLE 2: The parameters and calculations of different network models.

Network models	Parameters/ M	Calculations/GFLOPI
AlexNet	63.4	1.05
ResNet	13.47	5.24
SqueezeNet	3.63	0.91
Optimized SqueezeNet	3.80	0.72

optimized SqueezeNet are compared, whose results are shown in Table 2.

As shown in Table 2, AlexNet has the largest amount of parameters among the four networks, and the amount of calculations is at the middle level of the four networks. This is because AlexNet contains three fully connected layers, and the parameters in the three fully connected layers account for a relatively large amount, which results in a large amount of overall parameters of AlexNet. The amount of calculations is not very large because the number of layers in AlexNet is small and the parameters of the convolutional layer are small. The amount of parameters of ResNet is at the middle level of the four networks, but the amount of calculations is the largest. This is because the global average pooling layer is applied in ResNet, which effectively reduces the amount of structural parameters. The largest amount of calculation is due to the large number of layers in the ResNet. The amount of parameters in SqueezeNet is very small, and the amount of calculations is among the four networks. This is because the Fire module in SqueezeNet uses less 3×3 convolution, so its parameter is very small. The large amount of calculation is because there are more 1×1 convolutions in the Fire module. The optimized SqueezeNet introduces the idea of residual network and increases the network width, so the amount of parameters is slightly higher than that of SqueezeNet, but the amount of calculation is the smallest.

In summary, compared with several other network models, the optimized SqueezeNet model used in the paper has obvious advantages in both classification accuracy and recognition efficiency based on the Chinese painting sentiment classification. Therefore, it has certain research significance and practical value.

5. Conclusion and Outlook

As one of our country's traditional culture, Chinese painting embodies the long history and culture of our country and carries the thoughts, feelings, and humanistic spirit of the Chinese nation. Emotion can sublimate the artistic value of Chinese painting to a certain extent. Therefore, analyzing the emotions contained in Chinese paintings and assisting in the appreciation of art works are of great significance to the digital management of Chinese paintings and the promotion of the spirit of Chinese paintings. In this paper, combining the artistic characteristics of Chinese paintings and using a lightweight CNN, the sentiment characteristics of Chinese paintings are decomposed and analyzed in a quantified form, and a good sentiment classification result based on Chinese paintings is obtained. On the one hand, it can help the audi-

ence clearly understand the emotions and expressions of Chinese paintings, so as to better appreciate the art of Chinese paintings. On the other hand, it can directly display the regular pattern of Chinese paintings' sentiment, which is helpful for the digital management of Chinese paintings and the protection and dissemination of traditional Chinese culture and fully demonstrates the cultural heritage and spiritual outlook of the nation. However, in the future work, there are still the following issues that need to be further studied and improved.

Firstly, because the characteristics of Chinese paintings in different techniques or themes are very different, it is difficult to recognize their emotions with the same algorithm. In the future, we will study the sentiment recognition algorithms for multiple classes of Chinese paintings and perform more accurate calculations on the sentiment of Chinese paintings, so as to create greater value in the digital management and protection of Chinese paintings.

Secondly, we will further expand the database samples and establish a mathematical model based on the sentiment of Chinese paintings, so as to contribute to the future development of Chinese painting research.

Finally, in future work, virtual reality or augmented reality technology will be combined to improve the experience of visualizing the sentiment of traditional Chinese paintings, so as to promote the digital construction of museum cultural relics and strengthen the social and educational functions of digital museums.

Data Availability

The labeled data set used to support the findings of this study is available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was supported in part by the National Social Science Fund Project under grant 17BGL102, Excellent Project of Jiangsu Province Social Science Union under grant 15SYC-043, Soft Science Research of Wuxi Science and Technology Association under grant KX15-B-01, and Fundamental Research Funds for the Central Universities under grant 2015ZX18.

References

- [1] J. Sheng and J. Jiang, "Style-based classification of Chinese ink and wash paintings," *Optical Engineering*, vol. 52, no. 9, article 093101, 2013.
- [2] J. Sheng and J. Jiang, "Recognition of Chinese artists via windowed and entropy balanced fusion in classification of their authored ink and wash paintings (IWPs)," *Pattern Recognition*, vol. 47, no. 2, pp. 612–622, 2014.
- [3] M. Sun, D. Zhang, Z. Wang, J. Ren, and J. S. Jin, "Monte Carlo convex hull model for classification of traditional Chinese paintings," *Neurocomputing*, vol. 171, pp. 788–797, 2016.

- [4] P. Y. Chiang, C. V. Lin, and C. H. Tseng, "Generation of Chinese ink portraits by blending face photographs with Chinese ink paintings," *Journal of Visual Communication and Image Representation*, vol. 52, pp. 33–44, 2018.
- [5] M. Sun, Z. Dong, J. Ren, W. Zheng, and J. S. Jin, "Brushstroke based sparse hybrid convolutional neural networks for author classification of Chinese ink-washpaintings," in *2015 IEEE International Conference on Image Processing (ICIP)*, pp. 626–630, Quebec City, QC, Canada, 2015.
- [6] Z. Wang, D. Lu, D. Zhang, M. Sun, and Y. Zhou, "Fake modern Chinese painting identification based on spectral–spatial feature fusion on hyperspectral image," *Multidimensional Systems and Signal Processing*, vol. 27, no. 4, pp. 1031–1044, 2016.
- [7] R. Zhou, J. H. Han, H. S. Yang, W. Jeong, and Y. S. Moon, "Fast style transfer for Chinese traditional ink painting," in *2019 IEEE 9th International Conference on Electronics Information and Emergency Communication (ICEIEC)*, pp. 586–588, Beijing, China, 2019.
- [8] C. Zheng and Y. Zhang, "Two-stage color ink painting style transfer via convolution neural network," in *2018 15th International Symposium on Pervasive Systems, Algorithms and Networks (I-SPAN)*, pp. 193–200, Yichang, China, 2018.
- [9] W. Jiang, Z. Wang, J. S. Jin, Y. Han, and M. Sun, "DCT-CNN-based classification method for the Gongbi and Xieyi techniques of Chinese ink-wash paintings," *Neurocomputing*, vol. 330, pp. 280–286, 2019.
- [10] J. Sheng and Y. Li, "Classification of traditional Chinese paintings using a modified embedding algorithm," *Journal of Electronic Imaging*, vol. 28, no. 2, p. 1, 2019.
- [11] J. Sheng, C. Song, J. Wang, and Y. Han, "Convolutional neural network style transfer towards Chinese paintings," *IEEE Access*, vol. 7, pp. 163719–163728, 2019.
- [12] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 1031–1044, 2010.
- [13] Y. Sun and R. Fisher, "Object-based visual attention for computer vision," *Artificial Intelligence*, vol. 146, no. 1, pp. 77–123, 2003.
- [14] R. Datta, *Semantics and Aesthetics Inference for Image Search: Statistical Learning Approaches*, [M.S. thesis], The Pennsylvania State University. ProQuest Dissertations Publishing, 2009.
- [15] V. N. Cudivada, "A geometry-based representation for efficient and effective retrieval of images by spatial similarity," *IEEE Transactions on Knowledge and Data Engineering*, vol. 10, no. 3, pp. 504–512, 1998.
- [16] A. Hanjalic, "Video and image retrieval beyond the cognitive level: the needs and possibilities," in *Storage and Retrieval for Media Databases 2001*, vol. 2001, pp. 130–140, San Jose, CA, United States, 2001.
- [17] L. Marchesotti, N. Murray, and F. Perronnin, "Discovering beautiful attributes for aesthetic image analysis," *International Journal of Computer Vision*, vol. 113, no. 3, pp. 246–266, 2015.
- [18] T. Matthews, M. S. Nixon, and M. Niranjan, "Enriching Texture Analysis with Semantic Data," in *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1248–1255, Portland, OR, USA, 2013.
- [19] N. Mittal, D. Sharma, and M. L. Joshi, "Image Sentiment Analysis Using Deep Learning," in *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, pp. 684–687, Santiago, Chile, 2018.
- [20] P. Balouchian, M. Safaei, and H. Foroosh, "LUCFER: a large-scale context-sensitive image dataset for deep learning of visual emotions," in *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1645–1654, Waikoloa, HI, USA, 2019.
- [21] D. Kang, H. Shim, and K. Yoon, "A method for extracting emotion using colors comprise the painting image," *Multimedia Tools and Applications*, vol. 77, no. 4, pp. 4985–5002, 2018.
- [22] A. Yao, J. Shao, N. Ma, and Y. Chen, "Capturing Au-aware facial features and their latent relations for emotion recognition in the wild," in *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, pp. 451–458, New York, NY, USA, 2015.
- [23] D. Sen, S. Datta, and R. Balasubramanian, "Facial emotion classification using concatenated geometric and textural features," *Multimedia Tools and Applications*, vol. 78, no. 8, pp. 10287–10323, 2019.
- [24] Q. You, J. Luo, H. Jin, and J. Yang, "Robust image sentiment analysis using progressively trained and domain transferred deep networks," in *Twenty-ninth AAAI conference on artificial intelligence*, pp. 381–388, Austin, Texas, USA, 2015.
- [25] B. Jou and S. F. Chang, "Deep cross residual learning for multitask visual recognition," in *Proceedings of the 24th ACM international conference on Multimedia*, pp. 998–1007, New York, NY, USA, 2016.
- [26] V. Campos, B. Jou, and X. Giro-i-Nieto, "From pixels to sentiment: fine-tuning CNNs for visual sentiment prediction," *Image and Vision Computing*, vol. 65, pp. 15–22, 2017.
- [27] X. He and W. Zhang, "Emotion recognition by assisted learning with convolutional neural networks," *Neurocomputing*, vol. 291, pp. 187–194, 2018.
- [28] K. Song, T. Yao, Q. Ling, and T. Mei, "Boosting image sentiment analysis with visual attention," *Neurocomputing*, vol. 312, pp. 218–228, 2018.
- [29] Y. Fan, H. Yang, Z. Li, and S. Liu, "Predicting image emotion distribution by learning labels' correlation," *IEEE Access*, vol. 7, pp. 129997–130007, 2019.
- [30] T. Rao, X. Li, H. Zhang, and M. Xu, "Multi-level region-based convolutional neural network for image emotion classification," *Neurocomputing*, vol. 333, pp. 429–439, 2019.
- [31] W. Sun, H. Zhao, and Z. Jin, "A visual attention based ROI detection method for facial expression recognition," *Neurocomputing*, vol. 296, pp. 12–22, 2018.
- [32] X. Yang, Z. Wang, H. Deng et al., "Recognizing image semantic information through multi-feature fusion and SSAFE-based deep network," *Journal of Medical Systems*, vol. 44, no. 2, 2020.
- [33] X. Liu, N. Li, and Y. Xia, "Affective image classification by jointly using interpretable art features and semantic annotations," *Journal of Visual Communication and Image Representation*, vol. 58, pp. 576–588, 2019.
- [34] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size," 2016, <https://arxiv.org/abs/1602.07360>.
- [35] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Image net classification with deep convolutional neural networks," in *Neural Information Processing Systems Conference*, pp. 1097–1105, Barcelona, Spain, 2012.
- [36] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, Las Vegas, Nevada, USA, 2016.