Hindawi
Mobile Information Systems
Volume 2022, Article ID 8804711, 8 pages
https://doi.org/10.1155/2022/8804711

Research Article

A Multitask Convolutional Neural Network for Artwork Appreciation

Tian Tian[1] and Feng Nan[2]

[1]*Department of Product Design, College of Art and Design, Dalian Polytechnic University, Liaoning Province, China*
[2]*Network Information Center, Dalian Polytechnic University, Liaoning Province, China*

Correspondence should be addressed to Feng Nan; andynanfeng2021@163.com

Received 8 February 2022; Revised 28 February 2022; Accepted 8 March 2022; Published 14 April 2022

Academic Editor: Chia-Huei Wu

The computational aesthetics of pictorial art is an important part of human artistic creation, and the computational aesthetics of pictorial art images is a computationally computable human aesthetic process using machines, which has important applications and scientific significance in the automated analysis of large-scale paintings and the computational modeling of perception by machines. To this end, this paper proposes a multitask convolutional neural network model for emotion and rating of artworks. (1) An artwork appreciation dataset consisting of fifty Chinese paintings and fifty Western oil paintings was created, and twenty subjects were recruited to score the art appreciation of one hundred artworks in the dataset, covering both painting aesthetic evaluation and painting emotion evaluation. (2) Based on the artwork art appreciation dataset, an AlexNet-based convolutional neural network model is proposed to utilize the powerful feature extraction and classification capabilities of neural networks to complete artwork art appreciation, and an oversampling method and multitask learning method are used to improve the overall recognition accuracy. (3) Compared with the combination of traditional manual features + machine learning algorithms, the end-to-end multitask convolutional neural network proposed in this paper has the highest accuracy rate of 74.57%/71.43%/74.12%.

1. Introduction

Aesthetics is the study of the aesthetic scope of beauty and ugliness and the aesthetic activity of human beings, with art (especially painting) as the main object of study [1]. Although aesthetics is closely related to human sensibility, scientific methods such as psychology and neurology are also widely used in aesthetic research. In 1876, the German aesthete Fechner introduced experimental psychology into aesthetic research, using quantitative methods to interpret visual stimuli and objective measurements instead of subjective reasoning, creating experimental aesthetics [2]. In 2005, Computational aesthetics [3] was presented at the 1st Eurographics (EG) Conference on Computational Aesthetics of Graphics, Vision and Images. Workshop on Computational Aesthetics in Graphics, Visualization and Imaging, CAe 2005), the main meaning of which is to use machines to mimic human aesthetic processes, to autonomously perceive and cognize "beauty The main idea is to use machines to mimic human aesthetic processes, to perceive and cognize "beauty" autonomously, and to make aesthetic evaluations such as beauty and emotion. From experimental aesthetics to computational aesthetics, aesthetic research has evolved from using scientific methods to explain human aesthetic phenomena to using machines to imitate human aesthetic processes.

Earlier research in computational aesthetics focused on levels such as target recognition tasks (e.g., face recognition), which focused on the underlying visual features of objects and built bridges between the objective world and machine learning through statistical relationships between boundaries, shapes, and other art elements. With the popularity of target recognition applications, the industry's demand for image classification is no longer satisfied with basic target recognition but is turning to emotional and aesthetic aspects. Particularly with the development of artificial intelligence research, how to use computers to analyze and learn from subjective human aesthetic perception as an important auxiliary function of artificial intelligence has also attracted the attention of many researchers.

The goal of this paper is for the computer to describe, through appropriate models, the emotional reactions and subjective evaluations that arise when a person appreciates a work of art and categorizes them accordingly. Emotional semantics are more subjective than semantics at the cognitive level, and differences in a person's cognitive level, cultural background, and aesthetic standards can all impact emotional semantics. It is challenging to design feature extraction methods and perceptual classification algorithms based on established experience.

In recent years, deep learning methods based on convolutional neural networks have achieved significant success in many areas and are widely used in various fields. In the field of computer vision, the effect is immediate compared to manual feature extraction, such as classification problems [4, 5], object detection [6–8], and semantic segmentation of images [9]. This eliminates the limitations of manually extract ted features, especially when the image data containing multidimensional information can be directly input to the network, and effectively avoids the complexity of feature extraction during learning and data reconstruction during classification. The same neurons on the same mapping surface have the same weights, allowing the network to learn in parallel, reducing the computational effort, and shortening the learning cycle, which greatly improves the learning efficiency. For the complex problem of art appreciation, convolutional neural networks differ from manual feature extraction methods in that the latter tend to focus more on how to select the features extracted from an image to represent the emotion it is intended to reflect, whereas with convolutional neural networks, the algorithm itself focuses more on model building, which eliminates the limitations of manually extracted features and makes it possible to map between the underlying features and the higher-level semantics. This eliminates the limitations of manually extracted features and makes it possible to map between the underlying features and the higher-level semantics.

The main contributions of this paper are as follows.

(1) An artwork appreciation dataset comprising fifty Chinese paintings and fifty Western oil paintings was established, and twenty subjects were recruited to score the art appreciation of one hundred artworks in the dataset, covering both aesthetic evaluation and emotional evaluation of paintings

(2) Based on the art appreciation dataset, a convolutional neural network model based on AlexNet [4] was proposed, and the powerful feature extraction and classification ability of the neural network was utilized to complete the art appreciation of artworks

(3) Using the correlation between aesthetic evaluation labels and emotion labels, a multitask learning method is used to improve the overall recognition accuracy

(4) Compared with the combination of traditional manual features + machine learning algorithms, the end-to-end multitask convolutional neural network proposed in this paper has the highest accuracy rate of 74.57%/71.43%/74.12%.

2. Dataset

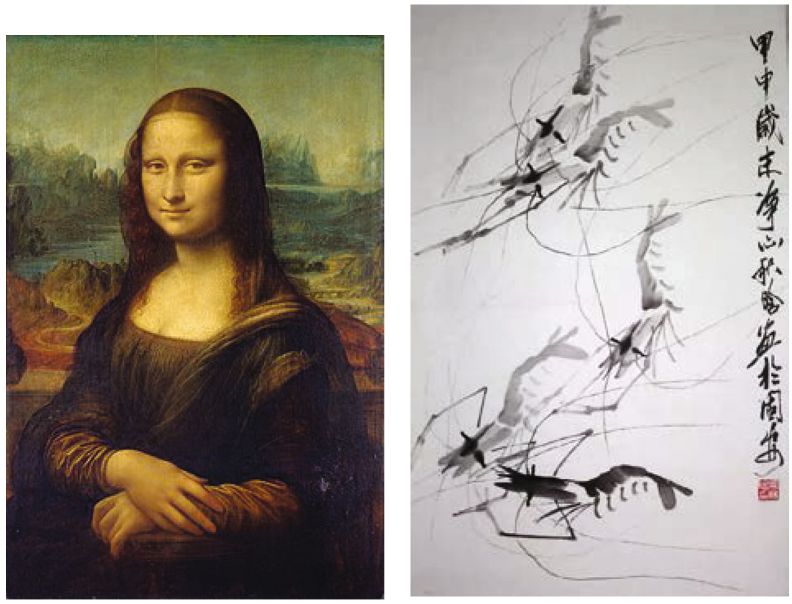
2.1. Artwork Selection. Fine artworks are sourced mainly from the Internet. The keywords "Chinese painting" and "oil painting" are used to search in the Internet, and certain representative artworks are selected by hand. These include landscape, figure, and bird and flower paintings in Chinese painting and figure, landscape, and still life paintings in oil painting. Figure 1 shows the classic paintings in the dataset.

2.2. Subjects. Twenty university students (10 males and 10 females) were recruited as annotators to label each artwork. University students are a group with a strong sense of self, as the development of one's sense of self is not only related to age and gender, but also to one's level of knowledge from the hierarchy. Generally the more literate a person is, the stronger their self-awareness is likely to be. In addition, the age and level of literacy of university students themselves makes them more aware of their own inner perception and analysis, trying to understand their own emotions and psychology, often observing social things in the outside world to the fullest extent, and having richer and deeper emotions. At the same time, because of the modern social level in which they live and their high cultural quality, university students are more receptive to new things and interested in participating in new scientific experiments and are also more likely to understand the requirements of the experiments and to make marked choices about the emotions of the paintings. Therefore, university students not only have a comprehensive psychological development, but also form a more complete self-awareness and have a stable personality, which is conducive to the completion of the questionnaire collection and the accuracy of the labels.

2.3. Questionnaire Design. The design of the questionnaire was completed using the visual analogue scale. A total of three questions were included, asking subjects to rate the emotional validity and arousal of the current artwork, as well as the overall aesthetics of the artwork. For emotional validity and arousal, a seven-point scale was used, and for aesthetics, a five-point scale was used, from low to high: very poor, poor, common, good, and very good. The questionnaire design is shown in Figure 2.

3. The End-to-End Multitask Convolutional Neural Networks for Art Appreciation of Artworks

3.1. Convolutional Neural Network. The pattern recognition algorithms in aesthetic computational models are support vector machines, linear discriminant, multilayer perceptron, artificial neural networks, etc., while artificial neural networks are the latest hot technology in machine learning



(a) (b)

Figure 1: Two typical artworks in our dataset.

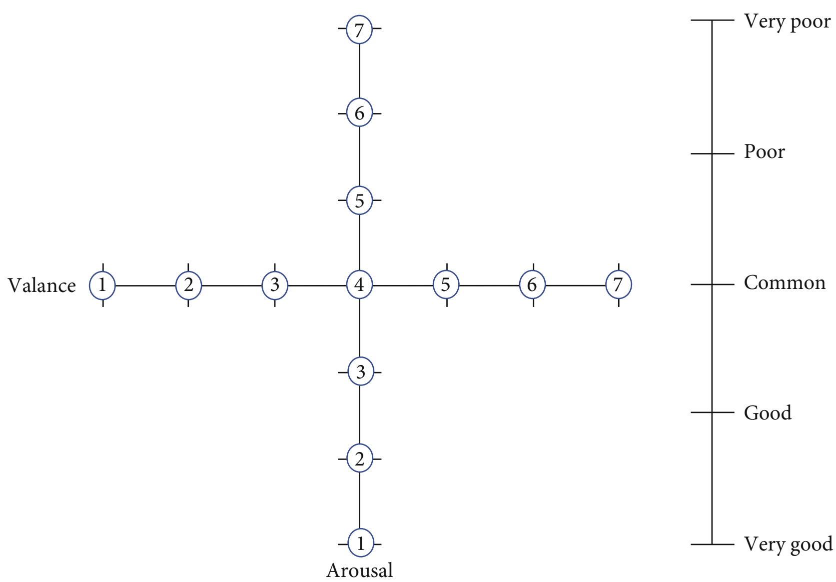


Figure 2: The questionnaire designment.

research. The model can ensure the objectivity, accuracy, and timeliness of the model. This chapter focuses on the improved convolutional neural network, a convolutional neural network classification model with a multitasking framework, and explores its application in art appreciation works.

Fukushima proposed Neocognitronl, which was able to simulate the way human eyes see and contributed to the birth of convolutional networks [10]. In 1988, Yann Le Cun proposed the Le Net-5 [11] network model to recognize handwritten numeric characters, which attracted much attention. In the following decades, various new types of convolutional neural networks have emerged. A convolutional neural network (CNN) consists of a multilayer network structure with input layers, convolutional layers, pooling layers, and fully connected layers, and generally a convolutional layer followed by a pooling layer, and finally several fully connected layers at the end of the network.

Convolutional neural networks are multilayer perceptions, but there are two key differences between MLPs and CNNs. First, in a CNN, the weights in the network are shared by the way the network performs a convolutional operation on the image. In this way, the model does not need to learn separate detectors for the same object that appears at different locations in the image, thus making the network isotropic with respect to the input. It also greatly reduces the number of parameters to be learned (i.e., the number of weights no longer depends on the size of the input image).

In the convolutional layer, the input image with a set of K kernels W = W1 W2, ⋯, WK and the added bias terms B = B1, B2, ⋯, BK.

The features are then nonlinearly transformed by $\sigma()$, and the same process is repeated for each convolution layer l.

$$X_k^l = \sigma\left(W_k^{l-1} * X^{l-1} + b_k^{l-1}\right). \quad (1)$$

The second key difference between CNNs and MLPs is the use of pooling layers in CNNs, where adjacent matrix elements are pooled using maximum or average pooling. This introduces a certain amount of translational invariance and again reduces the number of parameters in the network. The fully connected layers are usually added to the last layer of the convolutional layers. Similar to MLP, the output of the last layer is computed by the softmax function (Equation (2)) to generate the probabilities of the distribution of each class, and the network is trained using maximum likelihood.

$$P\left(y = \frac{j}{x}\right) = \frac{e^{x^T w_j}}{\sum_{k=1}^{K} e^{x^T w_k}}. \quad (2)$$

3.2. Multitask Learning. The key to multitask learning lies in finding the relationships between tasks. If the relationships between tasks are properly measured, then different tasks can provide each other with additional useful information, and using this additional information, better performing and more robust models can be trained. Multiple tasks share the same few hidden layers of the network between them, except that the network starts to fork to do different tasks in the network near the output layer. Different tasks learn some common features with low abstraction level by sharing several hidden layers at the bottom of the network, and the parameters shared at the bottom of this approach are identical. At the same time, each task is designed with its own task-specific layers to learn features at higher levels of abstraction, depending on the characteristics of each task. All tasks can share some relevant hidden layers while retaining task-specific output layers. This multitask learning approach can effectively reduce the risk of overfitting by averaging the noise. Moreover, the more related tasks there are, the less risk of overfitting the target task.

3.3. Oversampling. In order to make full use of the collected data to train the convolutional neural network model and to make the model more robust, the original image is modified to expand the dataset to train the model. In the literature [12], a method is proposed to expand the dataset to enhance the generalization ability of the network after learning, such as flipping or cropping. To verify the impact of this approach on the performance of the model for aesthetic appreciation tasks, we trained the convolutional neural network using several common expanded dataset approaches, including image cropping, flipping, image hue, brightness, saturation, and contrast. The model performance of the fine-tuned AlexNet model was retrained by expanding the ethnic art painting dataset in each of the above ways, and the combination of the ways that improved the model performance was selected to analyze the specific improvement of the model performance by the effective data expansion methods.

3.4. Model Structure. This model is adapted from the well-known neural network model AlexNet. The first layer is the convolutional layer. After convolving the image with a convolutional template of 11×11 and a sampling frequency of 4 pixels on three channels, the essential convolutional data is obtained. The output shape of first convolutional layer is $55 \times 55 \times 96$. Then a ReLU and Norm transformation is performed, followed by pooling to obtain a scale of $27 \times 27 \times 96$, which is passed to the next layer as the output. The second layer is still a convolutional layer, and the process is similar to the first layer, with convolution, ReLU, Norm, and pooling, and the output size is $13 \times 13 \times 256$. The convolutional layer process is the same, but the fourth and fifth layers differ from the first three layers in that the process is only convolution and ReLU, and the output size of the fourth layer is $13 \times 13 \times 384$. The output size of the fifth layer becomes $6 \times 6 \times 256$.

After five convolutional processes, the sixth layer enters the fully connected layer, which reaches 4,096 nodes after full connection. The seventh layer and eighth layer are similar to the sixth layer. And the final output is 1000 feature dimensions. As shown in Figure 3, the 1000 features are fed into different fully connected layers to obtain three model prediction outputs. Computing the results of the fully connected shared layer to obtain the sample prediction labels L^P, the multitask learning loss can be calculated according to the multiclass task label L of the artwork and L^P as follows:

$$\begin{aligned} loss &= Mean_{Absolute_{Error\left(L,L^P\right)}}, \\ Loss_{MTL} &= \sum_{i=1}^{i=3} loss_i + \alpha\|\theta\|. \end{aligned} \quad (3)$$

4. Results and Discussion

4.1. Feature Extraction

(1) Dark Channel Histogram. The dark channel is a concept referred to in classical image defogging algorithms [13]. Statistically, the dark channel is found to be gray in the cloudy parts of the image and can be used as a benchmark for image defogging. The dark channel is an alternative representation of the luminance feature and can be used in combination with the traditional luminance feature to improve the classification accuracy. The histogram of 10 bin was extracted from the dark channel of each image to obtain the 10-dimensional features D1 to D10.

(2) Boundary Complexity and Spatial Distribution. Edges are the intersection of regions with different attributes of an image, where the attributes of the region change abruptly

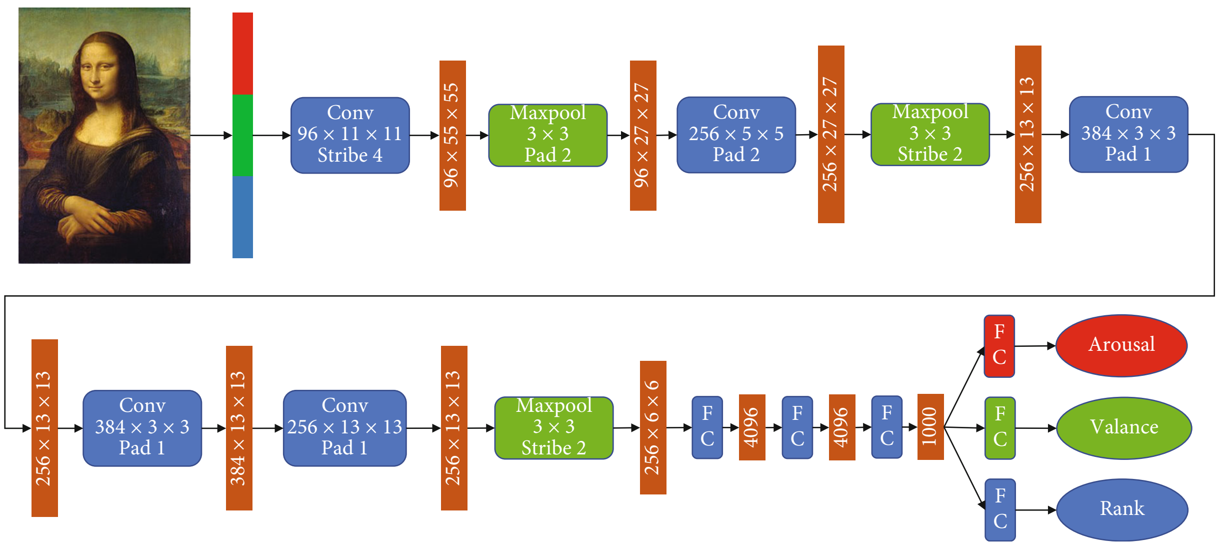


Figure 3: Structure of our proposed model.

Table 1: Different model results for arousal.

Models	Accuracy (%) Mean	Accuracy (%) Std
DT + features	52.51	2.30
SVM + features	55.35	2.18
RF + features	62.82	3.12
Model without MT	70.23	2.37
Our proposed model	74.57	1.91

Table 2: Different model results for valance.

Models	Accuracy (%) Mean	Accuracy (%) Std
DT + features	58.91	2.86
SVM + features	56.21	3.12
RF + features	62.12	2.82
Model without MT	68.12	2.56
Our proposed model	71.43	1.26

Table 3: Different model results for ratings.

Models	Accuracy(%) Mean	Accuracy(%) Std
DT + features	57.28	3.32
SVM + features	60.62	2.87
RF + features	65.09	4.11
Model without MT	71.67	3.17
Our proposed model	74.12	2.91

and contain a wealth of information. An edge image is a binary image, a two-dimensional Boolean matrix, obtained by determining the edge pixels of the original image. The ratio of edge pixel points to all pixel points is defined as the boundary complexity E. The original RGB image is transformed into a grayscale image, and the edges of the image are detected using the Canny operator [14]. The boundary complexity E1 to E5 was extracted by dividing the image into region (partitioning scheme from the literature [15]). The spatial distribution of the boundaries was also calculated by region finding ratio as

$$\begin{cases} E_6 = (E_1 + E_3)/(E_2 + E_4) \\ E_7 = (E_1 + E_2)/(E_3 + E_4) \\ E_8 = E_1/+E_4 \\ E_9 = E_2/+E_3 \end{cases}. \tag{4}$$

The symmetry character of the boundary complexity is thus constructed, together with the E10 for all regions of the image, a 10-dimensional boundary feature vector.

(3) HSV Color Histogram. The color histogram is a widely used color feature [16]. This method converts an RGB image into HSV space and uniformly quantizes the values of the H and V channels into 16 and 8, respectively, and performs the corresponding histogram statistics to obtain the histogram statistics of hue and luminance, H1 to H16 and V1 to V8, in 24 dimensions.

(4) Color Simplicity. The color conciseness feature was proposed in the literature [17]. The number of bins above this threshold is used as the color conciseness feature Hs.

(5) RGB Color Histogram and Luminance. The histogram quantifies the values of each of the three RGB channels into 10 uniformly and calculates the histograms R1 to R10, G1 to G10, and B1 to B10, for a total of 30 dimensions; the values of the grayscale image are quantified into 16 uniformly, and the histograms Gr1 to Gr16 are calculated, for a total of 16 dimensions.

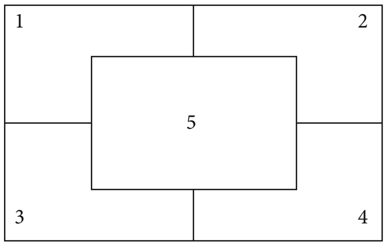


Figure 4: The region divisions.

(6) Number of Straight-Line Segments. The number of straight-line segments is calculated as the feature quantity L. The number of straight-line segments is calculated by using the Hough transform on the edge image.

(7) Neighborhood Difference Descriptor. The neighborhood difference descriptor calculates the similarity between the pixels around the boundary point and the boundary point itself, which is used to describe the sharpness of the Chinese brushwork and the gradation around the boundary point, similar to the sharpness in image processing. It is used in the literature [18] to describe the differences in painting techniques between brush painting and calligraphy, in 25 dimensions, noted as N1 to N25.

(8) Saliency Pattern. The saliency map is a model of visual attention proposed by Koch and Ullman [19]. The saliency map represents the visual attention of an image. This feature characterizes the distribution and strength of saliency in a picture, and the magnitude of S reflects the strength of saliency.

(9) Blur. Blur is a measure of image blur proposed by Ke et al. [17], which models the blur of an image based on the Gaussian blur hypothesis to calculate the degree of image blur. The image blur value is denoted as B.

(10) Contrast. The calculation of the image contrast characteristics is described in the literature [17]. The 256 bin histograms Hr, Hg, and Hb are calculated for each of the three RGB channels and added together to obtain Ha = Hr + Hg + Hb.

4.2. Machine Learning Algorithms

(1) Decision Tree. Decision tree (DT) is a decision analysis method that evaluates the risk of a project and judges its feasibility by forming a decision tree [20]. Based on the known likelihood of various scenarios occurring, it is necessary to find the probability that the expected value of the net present balue is greater than or equal to zero. It is called a decision tree because the decision branches are drawn as a graphical representation of the branches of a tree. In machine learning, a decision tree is a predictive model that represents a mapping between object attributes and object values. Entropy = the degree of clutter in the system, using algorithms ID3, C4.5, and C5.0 spanning tree algorithms using entropy. This metric is based on the concept of entropy in informatics theory.

(2) Support Vector Machine. Support vector machine (SVM) is a class of generalized linear classifier that performs binary classification of data in a supervised learning fashion, with a decision boundary of the maximum-margin hyperplane over the learned samples [21]. SVMs are sparse and robust classifiers that use a hinge loss function to compute empirical risk and add a regularization term to the solution system to optimize structural risk. SVM can perform nonlinear classification by kernel method, which is one of the common kernel learning methods.

(3) Random Forest. Random forest (RF) is an extended variant of the ensemble learning algorithm Bagging [22]. RF builds on the Bagging integration using a decision tree as the base learner and further adds random attribute selection to the training process of the decision tree. Specifically, a traditional decision tree selects an optimal attribute among all candidate attributes (assuming there are d) of the current node when choosing attributes for division; whereas in RF, for each node of the base decision tree, a subset containing k attributes is first randomly selected from the set of candidate attributes of that node, and then an optimal attribute is selected from this subset for division. The choice of the number of k attributes to be drawn is more important and is generally recommended. As a result, the "diversity" of the random forest base learner is not only perturbed by the samples, but also by the attributes, further enhancing the generalization of the final integration.

5. Results and Discussion

5.1. Algorithm Comparison Results. Ten-fold cross-validation was used to divide the training set into a test set and a training set to compare the performance of the combination of manual features with the three types of machine learning algorithms and the deep learning end-to-end network. In detail, the dataset is divided into ten parts, and nine of them are used in turn as training data and one as test data for the trials. Each trial yields a corresponding correct rate, and the mean and standard deviation of the correct rates of the 10 results are used as an estimate of the accuracy of the algorithm. The comparison results for effectiveness are shown in Table 1, for arousal in Table 2, and for rank in Table 3. It can be seen that although the random forest model has a high accuracy (62.82%/62.12%/65.09%) as an excellent integrated learning model, however, it is still inferior to our proposed end-to-end multitask convolutional neural network (74.57%/71.43%/74.12%), proving the superiority of our proposed method. In addition, the ablation experiments show that the multitask (MT) operation can improve the evaluation accuracy significantly.

5.2. Effect of Oversampling Method on Model Performance. This section compares the impact of expanding the dataset by cropping, flipping, and changing the hue, brightness, saturation, and contrast of the images on the performance of the model in the image sentiment classification task. In order to improve the generalization ability of the model, the

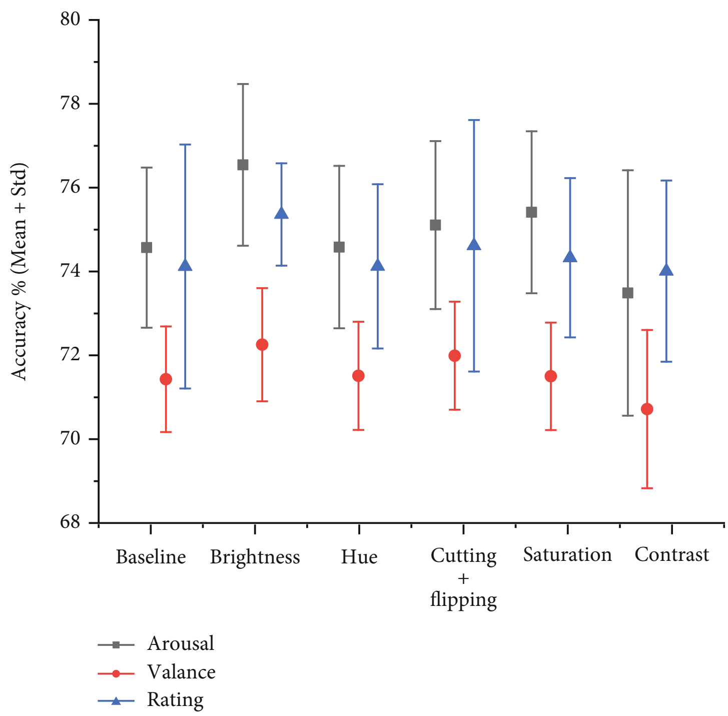


FIGURE 5: The comparison results of different oversampling methods.

dataset was changed in different ways as described in [18], and the changed data was fed back to the network to increase the learning ability of the model, which was oversampled as described in subsection 2.2. It can be seen from Figure 4 that the combination of crop + flip, luminance, and saturation can effectively improve the performance of the convolutional neural network model, not only in terms of accuracy, but also in terms of standard deviation. Changing the luminance is the most obvious way to improve the performance of the network, while changing the hue and contrast has a negative effect on the performance of the model, which is in line with our perception. Changing the hue means changing the color, which often gives intuitive psychological cues [31], e.g., red represents enthusiasm, happiness, and excitement, while its neighbor purple represents magic and weirdness. This is consistent with the findings of [11] that increasing or decreasing the contrast of an image can affect the model's attention, as shown in Figure 5.

6. Conclusion

In this study, an appreciation dataset of fine artworks including Chinese paintings and Western oil paintings was built, and subjects were recruited for sentiment and overall evaluation. A multitask convolutional neural network was trained to construct a fine art appreciation evaluation model, comparing traditional manual feature and machine learning methods to achieve the best results. In addition, an oversampling method was used to further improve the accuracy of the model.

Deep learning is currently developing at a rapid pace and has been widely used in many fields, but it has not been applied much in the appreciation of fine artworks. In this paper, we have made an attempt to appreciate fine artworks based on deep learning. However, there is still much work to be done, such as how to fuse certain prior knowledge of fine artworks into convolutional neural networks to uncover better features. It is believed that in the near future, the machine learning capability will be further enhanced, and the art appreciation capability will be significantly improved.

Data Availability

The labeled dataset used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares no competing interests.

Acknowledgments

This work is supported by the Dalian Polytechnic University.

References

[1] J. Shelley, *The Concept of the Aesthetic*, Stanford Encyclopedia of Philosophy, 2017.

[2] M. Skov and M. Nadal, "A farewell to art: aesthetics as a topic in psychology and neuroscience," *Perspectives on Psychological Science*, vol. 15, no. 3, pp. 630–642, 2020.

[3] F. Hoenig, "Defining computational aesthetics," in *Proceedings of the 1st Eurographics Conference on Computational*

Aesthetics in Graphics, Visualization and Imaging, pp. 13–18, Eurographics Association, Girona, Spain, 2005.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097–1105, 2012.

[5] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, http://arxiv.org/abs/1409.1556.

[6] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014.

[7] W. Liu, D. Anguelov, D. Erhan et al., "Ssd: Single Shot Multibox Detector," in *European Conference on Computer Vision*, Cham, 2016Springer.

[8] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016.

[9] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*Computer Vision Foundation.

[10] K. Fukushima, "Neocognitron: a hierarchical neural network capable of visual pattern recognition," *Neural Networks*, vol. 1, no. 2, pp. 119–130, 1988.

[11] X. Glorot and Y. Bengio, "Understanding the Difficulty of Training Deep Feedforward Neural Networks," in *Proceedings of the thirteenth international conference on artificial intelligence and statistics. JMLR Workshop and Conference Proceedings*, 2010.

[12] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: delving deep into convolutional nets," 2014, http://arxiv.org/abs/1405.3531.

[13] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2341–2353, 2011.

[14] B. Zitova and J. Flusser, "Image registration methods: a survey," *Image and Vision Computing*, vol. 21, no. 11, pp. 977–1000, 2003.

[15] Z. Dong and X. Tian, "Multi-level photo quality assessment with multi-view features," *Neurocomputing*, vol. 168, pp. 308–319, 2015.

[16] Y. Luo and X. Tang, "Photo and Video Quality Evaluation: Focusing on the Subject," in *European Conference on Computer Vision*, Berlin, Heidelberg, 2008Springer.

[17] Y. Ke, X. Tang, and F. Jing, "The design of high-level features for photo quality assessment," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 1IEEE.

[18] F. Gao, J. Nie, L. Huang, L. Y. Duan, and X. M. Li, "Traditional Chinese painting classification based on painting techniques," *Chinese Journal of Computers*, vol. 40, no. 12, pp. 2871–2882, 2017.

[19] L. Itti and C. Koch, "Computational modelling of visual attention," *Nature Reviews Neuroscience*, vol. 2, no. 3, pp. 194–203, 2001.

[20] A. J. Myles, R. N. Feudale, Y. Liu, N. A. Woody, and S. D. Brown, "An introduction to decision tree modeling," *Journal of Chemometrics: A Journal of the Chemometrics Society*, vol. 18, no. 6, pp. 275–285, 2004.

[21] Y. Lei, B. Yang, X. Jiang, F. Jia, N. Li, and A. K. Nandi, "Applications of machine learning to machine fault diagnosis: a review and roadmap," *Mechanical Systems and Signal Processing*, vol. 138, p. 106587, 2020.

[22] C. Zhang and Y. Ma, Eds., *Ensemble Machine Learning: Methods and Applications*, Springer Science & Business Media, 2012.