Hindawi
Advances in Multimedia
Volume 2022, Article ID 3799350, 11 pages
https://doi.org/10.1155/2022/3799350

Research Article

The Cultural Value Validity of Digital Media Art Based on Deep Learning Network Model

Yuan Ruan

Handan Polytechnic College, Handan, Hebei 056001, China

Correspondence should be addressed to Yuan Ruan; ruanyuan@hdvtc.edu.cn

Received 8 April 2022; Revised 10 May 2022; Accepted 16 May 2022; Published 28 May 2022

Academic Editor: Qiangyi Li

The development of Internet digital media technology has enabled more works of different artistic styles to be discovered, learned, and appreciated by art lovers. Artists generate new creative ideas from different artistic painting styles, resulting in many artistic creation styles that are mixed with different artistic creation techniques. However, in the face of an increasing number of digital media art works, the identification of artistic and cultural value is mainly done manually by professionals, which costs a lot of human and financial resources. Therefore, it is of great practical significance to study how to efficiently and accurately classify various types of artistic images to help users select images that meet their needs. In order to solve this problem, this paper proposes a deep learning neural network model based on a dual-core compression activation module. The convolution kernels of different sizes in one module are used to extract the overall features and local details of the image, and another module is used to achieve the main goal. The enhancement of features and the suppression of irrelevant features realize the evaluation of artistic and cultural value. The experimental results show that, compared with mainstream neural network models and traditional classification algorithms, the proposed algorithm has higher classification accuracy and higher recognition and classification efficiency.

1. Introduction

In the surging wave of technological innovation, the era of deep learning has come, and the current period is the best period for deep learning to develop rapidly and also the most controversial period. Deep learning has been developed for a short period of time, fused into the new era, and grafted onto and employed in professional fields such as intelligent interface, digital imaging, game design, business communication, for various types of cross-field innovation and thinking [1]. The breakthroughs in multimedia technology and big data technology have also made a good pavement for the development of deep learning.

Deep learning technology is widely used in image, text, and speech recognition; medical research; and other fields for its good autonomous learning ability and autonomous feature extraction [2]. Using machines to automatically learn and extract features of target objects is more objective and comprehensive than manual feature extraction and frees a lot of human and material resources from the task of feature extraction. Art images are different from ordinary shooting images in that there are similarities and differences in color, shape, texture, layout, etc. Using deep learning technology to learn and extract representative overall features and local detailed features of various art images to improve the accuracy of art image classification can help the general public to have a basic ability to distinguish between various art images, reduce the. reduce the misunderstanding. It is also important for the subsequent research of art image classification management [3].

Deep learning is faster, more comprehensive, and more efficient than traditional feature extraction methods, which has led a large number of scholars and enterprises to study deep learning, which is widely used in fine-grained image classification, multi-target object retrieval, face recognition, and other fields. These studies mainly focus on the visual processing of photographic images and the feature extraction of key content in images, while artistic images should not only be characterized according to the overall content, but also focus on the feature expression of local details. The

stylistic characteristics of an artistic image are distributed in the overall layout and local details. On this basis, some researchers manually extracted the underlying semantic information of art images and input the extracted features into convolutional neural networks for training and classification. divided the image into $4 * 4$ subblocks; extracted image semantic information such as color, texture, shape, and color layout of each subblock; input the extracted features into a neural network based on radial basis function for training; and then trained the image. The unique regions of the lines are extracted according to the Chinese ink stroke style, and the stroke features are extracted using a convolutional neural network to classify the ink artists [4]. Sheng et al. used edge detection to locate local parts, used histogram capture to reflect different stroke features and input them into parallel convolutional neural networks, and used the information entropy balance fusion method to make comprehensive decisions on the classification results of multiple neural networks, so as to realize the classification of painters [5]. Therefore, we study how to effectively and accurately classify various types of artistic images, help users to quickly filter the required images, and combine user preferences to make more accurate evaluations of the corresponding works. The current digital media art is very important and popular in usages. Therefore, this paper adopts a deep learning neural network model based on dual-core compression activation module and then uses another module to achieve the enhancement of main features and the suppression of irrelevant features to achieve the evaluation of artistic and cultural value, so as to improve the classification of digital media arts [6]. Accuracy, recognition ability, and classification efficiency save a lot of manpower and material resources in promoting the better development of digital media art [7].

2. State of the Art

In the 2015 3rd International Conference on Education Reform and Management Innovation (ERMI 2015), the authors made "Reform and Practice for Training Mode of 'Blending Art and Mechanics' Digital Media Talents with International Outlook" proposing the concept of reform and practice of "blending art and mechanics" digital media training mode with international perspective [8]. The 2015 International Conference on Arts, Design, and Contemporary Education (ICADCE 2015) presented "Practice and Reflection on the." In 2018, Jamerson Jeffrey published "Expressive Remix Therapy: Using Digital Media Art in Therapeutic Group Sessions with Children and Adolescents," which emphasized the thinking about teaching practices in digital media art [9]. The article discusses how to use digital media art for expressive blended therapies such as narrative therapy and expressive arts therapy to treat children and adolescents, guiding them on how to see and interact with the world [10].

Jie Deng published a paper in 2009 suggesting the need for the first mega-image database for computer vision researchers, and in 2010, the first ILSVRC—a large competition based on ImageNet, with an initial training sample of 1.2 million image clips—was held [11]. In terms of variety, the footage covered more than 1,000 categories and all had manual flags. After training, the program was evaluated on more than 50,000 test clips to determine whether it could classify image information clips [12].

In 2012, Professor Hinton led two graduate students to introduce the latest deep learning techniques into the ImageNet problem. They used a convolutional neural network model with eight layers, containing 650,000 neurons and 60 million free parameters. The team, represented by Professor Hinton, was able to train the program on 1.2 million images in almost six days with the help of two GPUs [13]. Based on this, 150,000 images were tested, and the model had an error rate of 15.3% in the first five categories of prediction. In the ImageNet competition conducted in 2012, the team achieved the first place test result among 30 groups. The Japanese team, which came second, had an error rate of 26.2%. This shows that neural networks are far ahead of other technologies in the field of image recognition and are expected to be a turning point for breakthroughs in digital technology [14].

Then Microsoft Research Asia, or MSRA, defended the crown of the 2015 ImageNet competition, increasing the depth of the network but reducing the learning efficiency. In order to solve the problem of decaying information validity in layer-by-layer transmission, the MSRA team tried to introduce the algorithm of "deep residual learning." The resulting MSRA deep residual learning model with 152 layers of neural networks set a new record in the first five categories tested, with an error rate of 3.57%, which is lower than the normal human error rate of about 5% [15]. Since 2017, the use of deep learning for image forensics has gradually increased. The literature proposes a combination of resampling features and SLTM to detect and locate image tampering operations. Subsequently, a dual-stream faster R-CNN network is proposed to detect tampered pictures. One branch of the two streams is the RGB stream, which is used to learn strong contrasting differences, tampering boundaries, etc. Another stream is used for noise extraction to find noise inconsistencies between real and tampered regions. Then, by bilinearly fusing the features from the two streams to enhance the tamper identification capability, the algorithm achieves better detection performance than each individual stream as well as the contrasting methods.

3. Methodology

3.1. The Basic Theory of Deep Learning. Artificial intelligence technology has developed rapidly in recent years, and deep learning technology has also developed rapidly with the computing power. Meanwhile, target detection is a fundamental problem in the field of machine vision as well as artificial intelligence, and its main goal is to pinpoint the class and location border information of various targets in images. Target detection methods have been widely used in various fields such as security surveillance, intelligent transportation, and image retrieval [16]. The research on target detection not only has a huge demand for applications in itself, but also provides the theoretical basis and research ideas for other machine vision tasks in related fields, such as

target tracking, face detection, pedestrian detection, and other techniques [17].

The supervised learning problem is to use training samples with labels to fit out the data with parameters w and b. In the simplest neural network, the input data x is processed by the neural network to output $h(x)$. If the neuron has three input values ($x1$), ($x2$), ($x3$) and an input intercept term +1 (generally written as b), the output is

$$\begin{aligned} h_{w,b}(x) &= f\left(W^T x\right) \\ &= f\left(\sum_{i=1}^{3} W_i x_i + b\right). \end{aligned} \tag{1}$$

In neural networks, nonlinearity can be obtained by activation functions: Sigmoid, ReLU, Tanh, etc.

A neural network is obtained by connecting numerous simple neurons together, and the output of one neuron can be used as the input of another neuron. Multilayer neural network models are shown in Figure 1. For neural network models, the ultimate goal is to get the final optimized parametric model for problems such as classification, detection, and segmentation [18]. Therefore, the most important part of the neural network model is the continuous optimization of its own model, and the most important algorithms to achieve this goal are the forward and backward propagation.

3.1.1. Forward Propagation Algorithm

$$\begin{aligned} z^2 &= W^1 x + b^1, \\ a^2 &= f\left(z^2\right), \\ z^3 &= W^2 a^2 + b^2, \\ h_{w,b}(x) &= a^3 \\ &= f\left(z^3\right). \end{aligned} \tag{2}$$

Equation (2) is the forward propagation step, where $(1) = x$ is the activation value of the input layer, and the activation value $(l+1)$ of the $l+1$st layer is calculated as shown in the following equation:

$$\begin{aligned} z^{l+1} &= W^l a^l + b^l, \\ a^{l+1} &= f\left(z^{l+1}\right). \end{aligned} \tag{3}$$

3.1.2. Backward Propagation Algorithm. The goal of the backpropagation algorithm optimization is to minimize the function $J(W, b)$ with W and b as parameters. To train the neural network, each $W_{ij}(l)$ parameter and each $b_i(l)$ parameter will be initialized to random values that are close to 0 and as small as possible; the main purpose of this is to make the symmetry invalid.

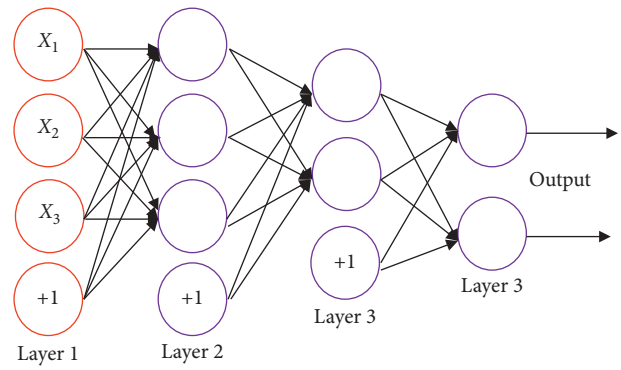

Figure 1: Multilayer neural network model.

To perform the forward propagation operation for (x), the activation values of all nodes in the neural network are calculated, and then the "residual" of each node i in the lth layer is calculated [18]. This residual value represents how much influence this node has on the residual of the whole output value and the final output node, which can directly get how much difference exists between the activation value obtained by the neural network and the actual comparison, as follows:

(a) Perform the forward propagation algorithm calculation to obtain the $L2$, $L3$, . . ., Ln activation values.

(b) The residuals of each output node i of the nlth layer can be calculated according to the following equation:

$$\begin{aligned} \delta_i^{(n_i)} &= \frac{\partial}{\partial z_i^{(n_i)}} J(W, b; x, y) \\ &= \frac{\partial}{\partial z_i^{(n_i)}} \frac{1}{2} \left\| y - h_{w,b}(x) \right\|^2 \\ &= \frac{\partial}{\partial z_i^{(n_i)}} \frac{1}{2} \sum_{j=1}^{S_{nj}} \left(y_j - a_j^{(n_j)} \right)^2 \\ &= \frac{\partial}{\partial z_i^{(n_i)}} \frac{1}{2} \sum_{j=1}^{S_{nj}} \left(y_j - f\left(z_j^{(n_j)} \right) \right)^2 \\ &= -\left(y_i - f\left(z_i^{(n_i)} \right) \right) f'\left(z_i^{(n_i)} \right) \\ &= -\left(y_i - a_i^{(n_i)} \right) f'\left(z_i^{(n_i)} \right). \end{aligned} \tag{4}$$

(c) $l = nl - 1, nl - 2, nl - 3, \ldots, 2$, when the residuals of the ith node in each layer can be calculated by the following equation:

$$
\begin{aligned}
\delta_i^{(n_i-1)} &= \frac{\partial}{\partial z_i^{(n_i-1)}} J(W,b;x,y) \\
&= \frac{\partial}{\partial z_i^{(n_i-1)}} \frac{1}{2}\left\| y - h_{w,b}(x)\right\|^2 \\
&= \frac{\partial}{\partial z_i^{(n_i-1)}} \frac{1}{2} \sum_{j=1}^{S_{nj}} \left(y_j - a_j^{(n_j-1)} \right)^2 \\
&= \frac{1}{2} \sum_{j=1}^{S_{nj}} \frac{\partial}{\partial z_i^{(n_i-1)}} \left(y_j - f\left(z_j^{(n_j-1)} \right) \right)^2 \\
&= -\sum_{j=1}^{S_{nj}} \left(y_j - f\left(z_j^{(n_j)} \right) \right) \frac{\partial}{\partial z_i^{(n_i-1)}} f\left(z_j^{(n_j)} \right) \\
&= -\sum_{i=1}^{S_{nj}} \left(y_j - f\left(z_j^{(n_j)} \right) \right) f'\left(z_j^{(n_j)} \right) \frac{\partial z_j^{(n_j)}}{\partial z_i^{(n_i-1)}} \\
&= \sum_{j=1}^{S_{nj}} \delta_j^{(n_j)} \frac{\partial z_j^{(n_j)}}{\partial z_i^{(n_i-1)}} \\
&= \sum_{j=1}^{S_{nj}} \left(\delta_j^{(nl)} \frac{\partial}{\partial z_i^{(n_i-1)}} \right) \sum_{k=1}^{S_{nj}-1} f\left(z_k^{n_i-1} \right) . W_{jk}^{nl-1} \\
&= \sum_{j=1}^{S_{nj}} \delta_j^{nl} W_{jk}^{nl-1} f'\left(z_i^{n_i-1} \right) \\
&= \left(\sum_{j=1}^{S_{nj}} W_{ji}^{\ nl-1} \delta_j^{\ nl} \right) f'\left(z_i^{\ n_i-1} \right).
\end{aligned}
\tag{5}
$$

By replacing the relationship between $nl-1$ and nl in (5) with the relationship between l and $l+1$, it can be deduced that

$$
\delta_i^{(l)} = \sum_{j=1}^{S_{l+1}} \left(W_{ji}^{(l)} \delta_i^{(l=1)} \right) f'\left(z_i^{(l)} \right). \tag{6}
$$

(d) Calculate the final required partial derivatives, as shown in the following equation:

$$
\begin{aligned}
\frac{\partial}{\partial W_{ij}^{(l)}} J(W,b;x,y) &= a_j^{(l)} \delta_i^{(l=1)}, \\
\frac{\partial}{\partial b_i^{(l)}} J(W,b;x,y) &= \delta_i^{(l=1)}.
\end{aligned}
\tag{7}
$$

A convolutional neural network is a feature extraction network that usually has multiple convolutional layers, as well as a pooling layer, and usually a fully connected layer at the end [19]. In a convolutional neural network, all weights are shared, and the pooling operation is shift-invariant. At the same time, convolutional neural networks have fewer parameters than multilayer neural networks and are very easy to train.

As shown in Figure 2, a convolutional neural network architecture usually includes an input layer, a multilevel convolutional layer, an activation layer, and an output part. However, in general, we can design different convolutional neural network architectures according to different task requirements, different hardware conditions, etc. The purpose of adding fully connected layers to a convolutional neural network structure is to predict the classification of a given object, but in practice, we can use 1×1 convolution instead of fully connected layers to reduce the number of parameters.

Feature extraction is an important step in the target detection algorithm. How to extract more effective information about the target object becomes the key to the target detection problem. The more valid the information, the higher the accuracy of detection. Furthermore, the quality of feature extraction is mainly determined by the performance of the designed neural network. Many more classical convolutional neural networks have appeared in recent years, which are described as follows.

3.1.3. AlexNet. The AlexNet network was proposed in the ILSVRC competition in 2012 and became famous for its performance on the ImageNet dataset, an image classification task, which far outperformed the then second-place finisher. AlexNet is formed by stacking simple convolutional layers. Five convolutional layers are used to extract feature information, and then 3 fully connected layers are used to process information to achieve classification.

AlexNet adopts a relatively deep neural network architecture to achieve feature extraction through the connection of convolutional layers, pooling layers, and fully connected layers. The ReLU activation function is widely used, and Dropout is used to avoid model overfitting. Dropout technology is equivalent to an ensemble algorithm. With a predetermined probability such as 0.5, some neurons are not involved in the calculation through Bernoulli's law during the training process of the network. By setting the output directly to 0, the outputs of all neurons are multiplied by the previously defined probability at test time. Dropout technology can effectively suppress the overfitting problem of the model and is widely used in fully connected layers and convolutional layers.

At the same time, AlexNet adopts Dropout method, and ReLU function adopts big data and multi-GPU training; all these strategies ensure that AlexNet obtains higher recognition rate than traditional image processing methods.

3.1.4. VGGNet. VGG model is another deep convolutional neural network designed in 2014 after AlexNet, which mainly enhances the feature extraction function of the network by increasing the depth of the network; it achieved

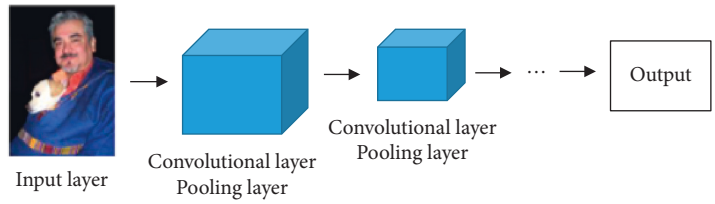

Figure 2: Convolutional neural network architecture.

the excellent result of runner-up in the ILSVRC competition. VGG has two different network structures, namely, VGG16 and VGG19.

The main difference compared to AlexNet is the use of small 3×3 convolutional kernels instead of large 5×5 and 11×11 kernels. The reason for this is that using smaller convolutional kernels enables the network to have fewer model parameters, while the perceptual field of two 3×3 kernels is the same as that of one 5×5 kernel, but with one more convolutional layer, which increases the effect of nonlinearity and enables the extraction of more detailed feature information. This increases the effect of nonlinearity and enables the extraction of more detailed feature information.

3.1.5. ResNet. ResNet was proposed because there is a problem in deep convolutional neural networks: the deeper the network model is, the harder it is to train it, mainly because there are two problems in deep networks: first, gradient disappearance and explosion; second, degradation problem. The gradient disappearance and explosion problems can usually be solved by batch normalization (BN) techniques, but how to solve the degradation problem? This is the reason why ResNet proposes a deep residual structure and experimentally proves its feasibility to ensure that the network does not degenerate during the stacking process. The residual structure is shown in Figure 3.

As can be seen in Figure 3, the residual structure has a channel mapping directly to the output, so that the network can learn the residuals of the input and output, so that the gradient disappearance problem does not occur during backpropagation, making a deeper network structure possible.

In 2014, Li et al. proposed the global mean pooling operation, which is similar to the mean pooling operation, except that the sliding window size of global mean pooling is the same as the width and height of the feature map to be manipulated. Unlike the fully connected layer, the global mean pooling operation sums the values of the two-dimensional feature layer output from the convolutional or pooling layer and then averages them to obtain one feature value for each feature layer; i.e., the feature map of size $W * H * D$ becomes a $1 * 1 * D$ size tensor, so that each feature layer can obtain one feature value after the global mean pooling process. Compared with the fully connected layer, the redundant parameters and overfitting problems are reduced, and the expression formula is as follows:

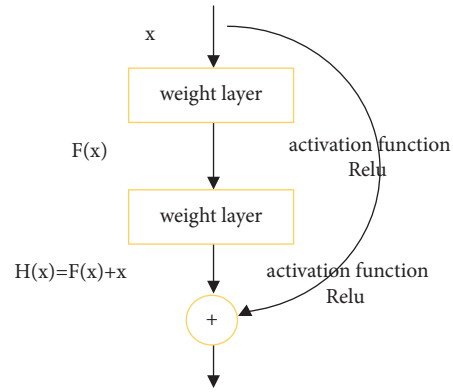

Figure 3: ResNet residual structure.

$$S_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} U_z(i, j), \tag{8}$$

where U_c denotes the cth layer of the feature map U and S_c denotes the feature value obtained after the mean processing of U_c. The global mean pooling process is shown in Figures 4 and 5.

3.2. Digital Media Art. Compared to other arts, digital media art differs mainly in the digital technologies and methods that must be used when designing a digital media art work or when presenting the final result of the work.

Digital media art is a kind of art using digital media as a medium, and from the perspective of disciplinary composition, digital media art involves the following disciplines: (1) design, (2) media technology, (3) visual arts, and (4) computer graphics. Most of these disciplines are interrelated, and most of them are expressed in the form of digital media, with artworks or design products as the final expression content; for example, digital media art is expressed in some multimedia web pages and interactive installation art. Digital media art is also very special in the form of media dissemination—with the chosen dissemination platforms being the Internet, cell phones, and electronic interactive media—and its dissemination forms are also diverse.

Visual art and design are part of the artistic layer of digital media art. Unlike other jobs, staff engaged in digital media art practice should have certain imagination. Digital media art practice is a process of discovering beauty and creating beauty. Designers have to create works that meet the esthetic interests and esthetic experiences of the public according to people's esthetic laws, in order to achieve an increase in esthetic purpose and emotional resonance.

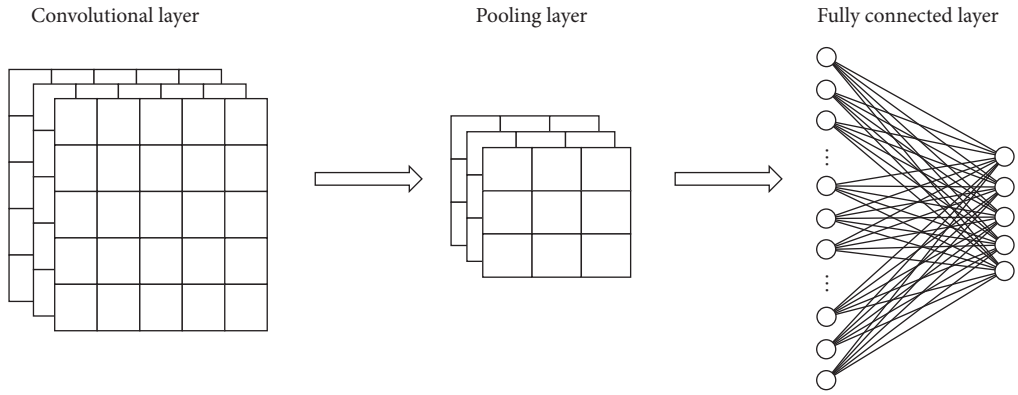

Figure 4: Input of fully connected layer.

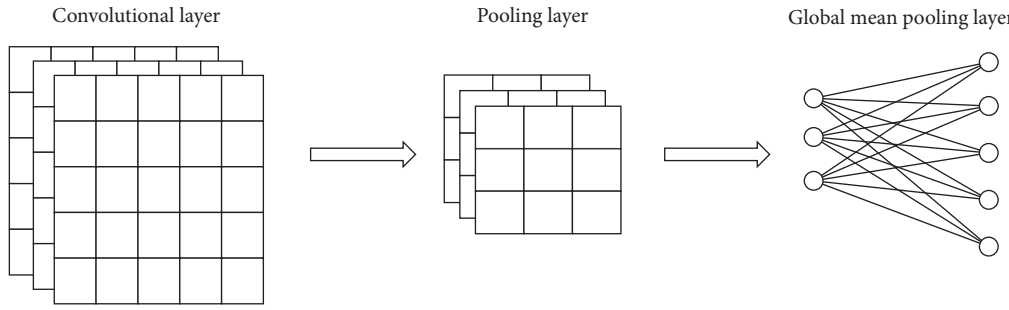

Figure 5: Global mean pooling input.

Digital media also has a technological layer. The technology layer speaks of the technical aspects of digital media, which are supported by computer special effects, computer technology, multimedia technology, etc. Under the support of these technologies, digital media technology has the characteristics of interactivity, integration, and immersion. Media arts cover the Internet and sociology, which are able to be integrated for media considerations, such as public media, mass media, and online media.

From the application level, digital media arts are obvious arts that include social information service functions, including digital entertainment, e-government, and e-commerce. The application level is more closely related to traditional industries, such as layout design, stage modeling, image planning, product design, and clothing design.

The composition of digital media art is a pyramid structure; it has four sides, representing the technical level, the artistic level, the media level, and the service level, each of which exists by virtue of the other levels, being interdependent and indispensable and running to construct the overall structure of the pyramid. Accordingly, the content of digital media art is not simple; not only does it refer to a discipline, but it also represents the art forms and service models that will emerge in the future life.

4. Result Analysis and Discussion

4.1. Dual-Kernel Compressed Activation Module Design. In recent years, convolutional neural networks have achieved good results in extracting overall features and local detail features of images. The large convolutional kernel in Inception-v4 module can extract overall features of images, and the small convolutional kernel realizes the extraction of local detail features. However, these modules mainly perform the transformation of spatial information, despite the dependencies between the function channels, and do not further improve the extracted functions. The SE (Squeeze and Excitation) module proposed by Hu et al. in 2017 incorporates the dependencies between the spatial dimensions of the feature map and the channels, reinforces the useful features in the feature map, and suppresses the less useful features. The schematic diagram of the SE module is shown in Figure 6. The SK (Selective Kernel) module proposed by Li et al. can adaptively adjust the perceptual field size to extract the overall features and local detail features of the input sample. The SK module is schematically shown in Figure 6. According to the characteristics of SE and SK modules, this paper constructs a neural network model based on Double Kernel Squeeze and Excitation (DKSE), as shown in the figure. The SE module is mainly used to extract the main features, and the SK module is mainly used to flexibly select the size of the receptive field according to the different digital art pictures, so as to realize the automatic adjustment of the convolution kernel.

The DKSE module performs feature fusion, feature compression, and activation on the fused feature maps; then performs weighted mapping of the processed features to the feature maps on each branch; and finally performs fusion of the corresponding elements on the mapped feature maps to enhance the role of key features in the overall and local detail features. The overall process is shown in Figure 7.

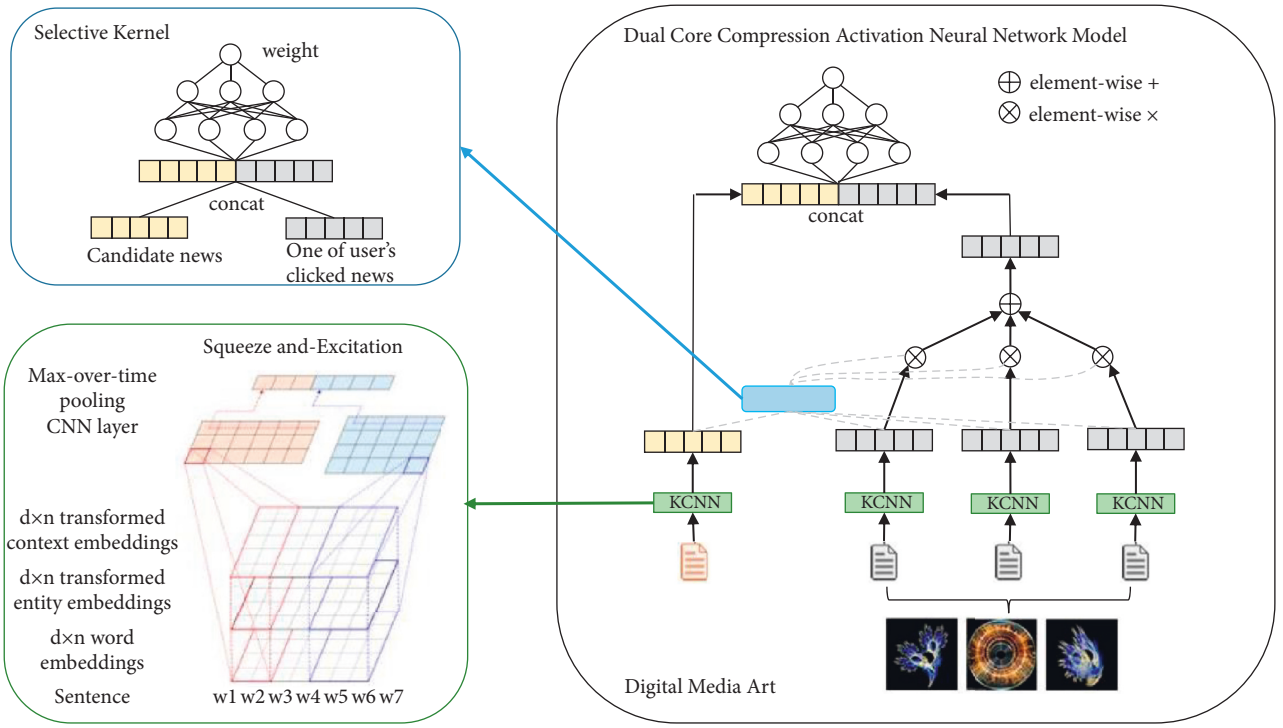

Figure 6: Schematic diagram of algorithm framework structure.

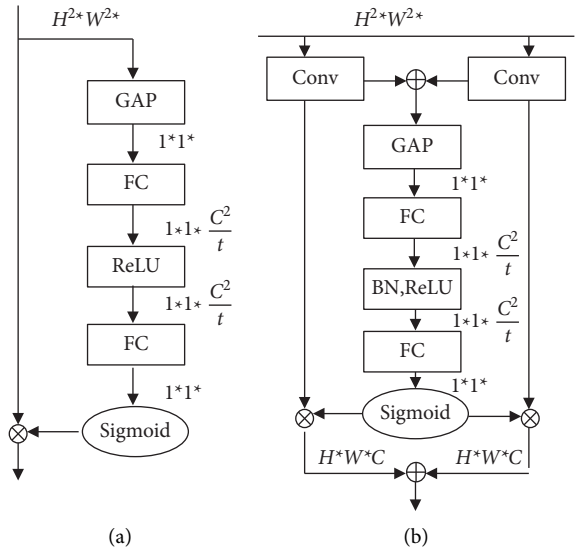

Figure 7: (a) SE module and (b) SK module.

The proposed DKSE module combines the features of the SE module and the SK module, which can better enhance the overall style features and local detail features of the extracted artistic images. The added processing flow is shown in Figure 8.

$$V = \sum_{i=1}^{n} U_i \cdot F_{ex}\left(F_{sq}\left(F_{jp}(U)\right)\right). \tag{9}$$

4.2. Deeply Separable Convolution. Depthwise separable convolution improves the one-step calculation method in traditional convolution and splits it into a superposition of depthwise convolution and 1×1 convolution in a cascaded manner, realizing the traditional decoupled operations for convolution. Although traditional convolution can achieve excellent feature extraction effect, it has an obvious problem; that is, the amount of parameters and calculation is too large, and there are a large number of similar features in the extracted feature information. The traditional convolution process considers all channels of the corresponding image region simultaneously, as shown in Figure 8. In the process of depth separation convolution, on the other hand, considering the spatial regions and channels of the

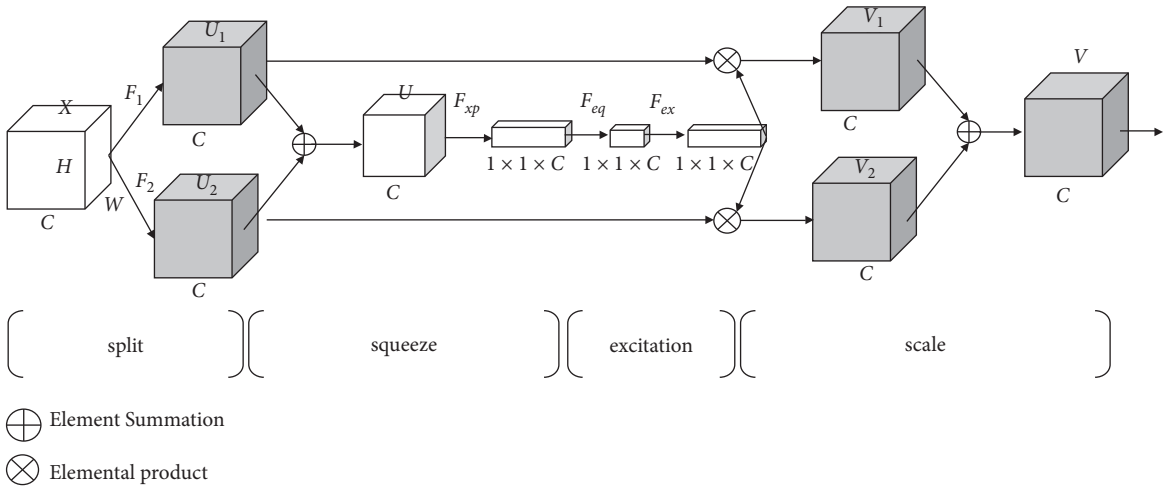

Figure 8: DKSE module.

corresponding image, the convolution process can be divided into depthwise convolution and point convolution operations to increase network depth and reduce computational effort. The convolution process is as follows (Figure 8).

For the input data $X \in RH * W * C$, the convolution operation is performed with C′ convolution kernel filters to obtain the feature map $Y \in RM * N * C$. H, W, and C denote the height, width, and number of channels of the input samples, respectively; h and w denote the height and width of the convolution kernels; and M and N denote the height and width of the output samples, respectively. The number of parameters of the convolution kernels and the amount of computation in the traditional convolution process are as follows:

$$\begin{aligned} S_c &= h \times w \times C \times C', \\ C_c &= M \times N \times S_c. \end{aligned} \tag{10}$$

The number of parameters of the convolution kernel for the depth-separable convolution operation is obtained by summing the number of parameters of the convolution kernel in the depth convolution and point-by-point convolution operations, and the depth convolution is mainly used for the feature map dimensionality reduction, while the point-by-point convolution is mainly used for the feature map channel expansion. The number of convolution kernel parameters and the computation volume of the deep separable convolution operation are given by the following equation:

$$\begin{aligned} S_d &= h \times w \times C + 1 \times 1 \times C \times C', \\ C_d &= M \times N \times S_d. \end{aligned} \tag{11}$$

Based on the above equation, the ratio of the ordinary convolution to the depth-separable convolution in terms of the number of convolution kernel parameters can be calculated as follows:

$$\begin{aligned} \frac{S_d}{S_c} &= \frac{C_d}{C_c} \\ &= \frac{1}{hw} + \frac{1}{C'}. \end{aligned} \tag{12}$$

In summary, it can be concluded that the depth-separable convolution operation can effectively reduce the amount of network computation and the number of network model parameters.

4.3. Building a Dual-Core Compressive Activation Neural Network Model. In this paper, we mainly use the depth-separable convolution and DKSE modules to build a convolutional neural network. The first layer of the network uses the null convolution for feature extraction of the original art image, and the null convolution has a larger network perceptual field compared with the normal convolution, which can keep more internal data structure and original image information without increasing the computational effort, as shown in Figure 9. Figure 9 represents the conventional convolution process with a fill of 0, a step size of 1, and a convolution kernel size of $3 * 3$, while Figure 9 shows a $3 * 3$ convolution kernel with an expansion rate of 2 and a perceptual field of $5 * 5$, which can preserve more original information during the process of null convolution. The depth-separable convolution is composed of depth convolution and point-by-point convolution. The DKSE module embedded in the depth-separable convolution operates with the following equation:

$$Y(X) = \left(Y_1 * Y_2 * Y_3\right)(X). \tag{13}$$

In Figure 10, we present the case of taking 3×3 and 5×5 ordinary convolution operations on DKSE branches. To reduce the overfitting phenomenon during the training process, we use $L2$ regularization method in the point-by-point convolution operation and Dropout processing before

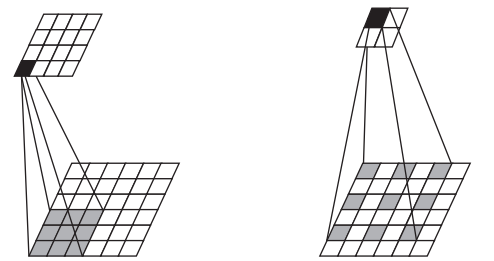

Figure 9: (a) Original convolution operation and (b) null convolution operation.

ID	Input size	Convolution type	Convolution kernel size
1	299x299x3	Dilated_conv/s1	3x3, 32, d=2
2	299x299x3	Depthwise_conv/s3	3x3
3	2	Dep_Separable_conv/s1	3x3, 64
4	99x99x32	Dep_Separable_conv/s2	3x3, 128
5	99x99x64	Dep_Separable_conv/s1	3x3, 128
6	50x50x128	DKSE_block/s1	3x3, 5x5, 128, r=4
7	50x50x128	Dep_Separable_conv/s2	3x3, 256
8	50x50x128	Dep_Separable_conv/s1	3x3, 256
9	25x25x256	DKSE_block/s1	3x3, 5x5, 256, r=4
10	25x25x256	Dep_Separable_conv/s2	3x3, 512
11	25x25x256	Dep_Separable_conv/s1	3x3, 512
12	13x13x512	DKSE_block/s1	3x3, 5x5, 512, r=4
13	13x13x512	Dep_Separable_conv/s2	3x3, 3x3, 1024
14	13x13x512	Dep_Separable_conv/s1	3x3, 1024
15	7x7x1024	Global Average Pool	7x7
16	7x7x1024	Fully Connected	1024x5
17	1x1x1024	Softmax	-
	1x1x5		

Figure 10: Dual-core compressed activation neural network model.

and after the final global mean pooling process. The second layer of deep convolution operation is followed by batch normalization and ReLU function processing. Dilated_ conv denotes a null convolution layer with convolution kernel size of 3×3 and expansion rate of $d = 2$, $s1$ denotes a convolution step of 1, $s2$ denotes a convolution step of 2, Depthwise_conv is a deep convolution operation, and Dep_Separable_ conv denotes the depth-separable convolution layer.

4.4. Effect of Parameters of the Dual-Core Compression Activation Module

(1) Position. In order to see the effect of DKSE module on the classification of art images at different positions of the network model, we place the DKSE module alone at the positions of network model ID numbers 6, 9, and 12, respectively, and take 3×3 and 5×5 convolutional kernels on the branches of DKSE module, and the descent rate r value is taken as 4. As shown in Table 1, when the DKSE module alone is placed at the position of network model ID 6, the classification of art images has the highest classification accuracy and consumes the least amount of computation. However, it can also be found that the deeper the DKSE module is placed in the network model, the more the parameters used for network training are. This is mainly because the deeper the model is placed, the more the number of channels of the feature map increases; in addition, the number of convolution kernels tends to increase.

Table 1: Accuracy of DKSE module at IDs 6, 9, and 12.

ID	Parameters (M)	Time (min)	Accuracy (%)	GFLOPs
6	2.7	1800	87.24	0.75
9	4.4	2043	87.14	1.77
12	11.1	1860	86.46	3.37

(2) Descent rate r and branch convolution kernel size. The descent rate r and branch convolution kernel

Table 2: Comparison of r value and convolution kernel size classification results.

r	1×1	3×3	5×5	Parameters (M)	Time (min)	Accuracy (%)
	√	√		2.3	1367	87.26
	√		√	2.6	1680	87.58
4		√	√	2.7	1800	87.24
	√	√	√	2.7	2370	87.35
	√	√		2.3	1230	86.35
	√		√	2.6	1220	86.85
16		√	√	2.7	1440	86.15
	√	√	√	2.7	1770	86.42

Table 3: Classification results of DKSE module branching taking different null convolution kernels.

r	K3	K5	K7	Parameters (M)	Time (min)	Accuracy (%)
	√	√		2.4	1530	86.07
	√		√	2.4	1657	86.00
4		√	√	2.4	1560	84.50
	√	√	√	2.6	2220	86.40

size are an important set of parameters in the DKSE module that control the computational resources and experimental accuracy. We take the network model of the DKSE module placed at ID number 6 alone and experiment and analyze the r value and branch convolution kernel size in the DKSE module. As shown in Table 2, when the size of the convolution kernel on the branch of the DKSE module is fixed, the classification results are higher when the r value is taken as 4 than when the r value is taken as 16; when the r value is fixed, the two-branch DKSE module takes shorter training time and fewer parameters than the three-branch DKSE module, and the experimental results have the highest accuracy when the convolution kernel on the branch of the DKSE module is taken as 1×1 and 5×5, respectively.

(3) Null convolution. In order to compare the effect of the null convolution kernel on the artistic image feature extraction, we take different sizes of the null convolution kernel on the DKSE module to conduct experiments on the data of this paper, as shown in Table 3, where K3 denotes the ordinary 3×3 convolution kernel; K5 denotes the 3×3 convolution kernel with the expansion rate of 2, and the perceptual field of its null convolution kernel is 5×5; and K7 denotes the 3×3 convolution kernel with the expansion rate of 3, and the perceptual field of its null convolution kernel is 7×7. The experimental results show that the parameters of the null convolution are fewer than those of the normal convolution with the same field, but the classification accuracy is not as high as that of the normal convolution because the combination and distribution of the features of the art images are random, and the null convolution may miss some important features in the process of feature extraction, which affects the classification results of the model.

5. Conclusion

In this paper, we introduce the basic components of convolutional neural network and its working principle, and based on the understanding and mastery, we propose the dual-core compressed activation module (DSKE) to extract the overall features and local detail features of art images according to the working principle between network modules and the characteristics between art images. The art images are processed for data enhancement, and then the convolutional neural network is built with the DKSE module and several depth-separable convolutions to achieve effective classification of five types of art images.

The experimental validation of the proposed dual-core compressed activation neural network-based art image classification algorithm is carried out in this paper. The experimental validation results show that data enhancement processing of samples can effectively improve the classification accuracy. Compared with mainstream neural network models and traditional classification algorithms, the algorithm in this paper has a higher classification accuracy, verifies the influence of the parameters of the dual-core compressed activation module on the classification of the model, and obtains a reasonable set of module configuration parameters. This paper uses the Grad-CAM algorithm to visualize and analyze the regions of the network model in the learning process that depend on accuracy. The analysis results show that the network of this paper extracts the overall features and local detail features of art images better. The classification of class 3, class 4, class 5, and class 6 images using the network of this paper shows that the classification accuracy does not improve due to the reduction of sources. The classification accuracy of the network model does not

change much when the dual-core compressed activation module is improved, but the training time can be improved.

Data Availability

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

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the Handan Polytechnic College.

References

[1] A. Pratapa, M. Doron, and J. C. Caicedo, "Image-based cell phenotyping with deep learning," *Current Opinion in Chemical Biology*, vol. 65, pp. 9–17, 2021.

[2] P. Ma, L. Chun Pong, Yu Ning et al., "Image-based nutrient estimation for Chinese dishes using deep learning," *Food Research International*, vol. 147, 2021.

[3] Z. Bai and X.-L. Zhang, "Speaker recognition based on deep learning: an overview," *Neural Networks*, vol. 140, 2021.

[4] Y. Qiu and J. Lu, "A visualization algorithm for medical big data based on deep learning," *Measurement*, vol. 183, Article ID 109808, 2021.

[5] J. Guan, Y. Xu, L. Ding, X. Cheng, C. S. Lee Vincent, and C. Jin, "Automated pixel-level pavement distress detection based on stereo vision and deep learning," *Automation in Construction*, vol. 129, 2021.

[6] B. Murray and P. L. Prasad, "An AIS-based deep learning framework for regional ship behavior prediction," *Reliability Engineering & System Safety*, vol. 215, 2021.

[7] S. Wang, X. Gu, S. Luan, and M. Zhao, "Resilience analysis of interdependent critical infrastructure systems considering deep learning and network theory," *International Journal of Critical Infrastructure Protection*, vol. 35, 2021.

[8] M. Li, L. Han, J. Chen et al., "swFLOW: a large-scale distributed framework for deep learning on Sunway TaihuLight supercomputer," *Information Sciences*, vol. 570, 2021.

[9] A. W. Ryan and A. Jørund, "Digital storytelling, student engagement and deep learning in Geography," *Journal of Geography in Higher Education*, vol. 45, no. 3, 2021.

[10] Y. Zhu, T. Brettin, F. Xia et al., "Publisher Correction: converting tabular data into images for deep learning with convolutional neural networks," *Scientific Reports*, vol. 11, no. 1, Article ID 14036, 2021.

[11] Z. Liu, L. Jin, J. Chen et al., "A survey on applications of deep learning in microscopy image analysis," *Computers in Biology and Medicine*, vol. 134, 2021.

[12] C. A. Neves, E. D. Tran, N. H. Blevins, and P. H. Hwang, "Deep learning automated segmentation of middle skull-base structures for enhanced navigation," *International Forum of Allergy & Rhinology*, vol. 11, no. 12, pp. 1694–1697, 2021.

[13] C. Lu, Z. Liu, X. Wen, L. Cao, H. Wu, and S. Qin, "Resonance calculation based on the deep learning method for treating the non-uniform fuel temperature distribution in PWRs," *Annals of Nuclear Energy*, vol. 160, 2021.

[14] C. Shen, C. Wang, M. Huang, Xu Ning, S. Van Der Zwaag, and W. Xu, "A generic high-throughput microstructure classification and quantification method for regular SEM images of complex steel microstructures combining EBSD labeling and deep learning," *Journal of Materials Science & Technology*, vol. 93, 2021.

[15] K. Sebastian, K. Nils, K. H. U. Syed, H. V. Manuel, N. Fabian, and L. Markus, "Towards scalable economic photovoltaic potential analysis using aerial images and deep learning," *Energies*, vol. 14, no. 13, 2021.

[16] N. Tits, "Controlling the emotional expressiveness of synthetic speech: a deep learning approach," *4OR*, vol. 20, 2021.

[17] Y. Kim, H. J. Lee, and J. Shim, "Developing data-conscious deep learning models for product classification," *Applied Sciences*, vol. 11, no. 12, p. 5694, 2021.

[18] A. A. Salih, S. Y. Ameen, S. R. M. Zeebaree et al., "Deep learning approaches for intrusion detection," *Asian Journal of Research in Computer Science*, 2021.

[19] P. G. Fernando, S. Rachel, and O. Sébastien, "TorchIO: a Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning," *Computer Methods and Programs in Biomedicine*, vol. 208, 2021.