IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received May 26, 2020, accepted June 13, 2020, date of publication July 10, 2020, date of current version July 30, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3007353

Image Recommendation Algorithm Based on Deep Learning

PEI YIN[1] AND LIANG ZHANG[2]
[1]Business School, University of Shanghai for Science and Technology, Shanghai 200093, China
[2]School of Aerospace Engineering, Tsinghua University, Beijing 100084, China

Corresponding author: Liang Zhang (zl19@tsinghua.org.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 71601119, Grant 71771177, and Grant 71871143, in part by the Chenguang Program supported by Shanghai Education Development Foundation and Shanghai Municipal Education Commission under Grant 16cg53, and in part by the Fundamental Research Funds for the Central Universities of China.

ABSTRACT With the rapid development of network technology, the number of digital images is growing at an alarming rate, people's demand for information gradually shift from text into images. However, it is very difficult for users to quickly find the images they are interested in from the large number of image libraries. The purpose of this paper is to study the image recommendation algorithm based on deep learning. In this paper, image classification algorithm is firstly studied. LReLU - Softplus activation function is formed by combining LReLU function and Softplus function, and CNN is improved. Then, an image retrieval model based on local sensitive hash algorithm is proposed in this paper. This model calculates the distance in hamming space for the binary hash code generated by mapping. Euclidean distance is calculated inside the result set after similarity measurement to improve the accuracy, and the image retrieval model is constructed. Finally, an image recommendation model based on implicit support vector machine (SVM) is proposed in this paper. This image recommendation method combines image text information and image content information. The experimental results show that the proposed image recommendation model can meet the practical needs. In this paper, the overlap rate between the CNN-based recommendation model and the human recommendation algorithm was tested, and the coincidence degree of the two recommended images reached 88%.

INDEX TERMS Deep Learning, convolutional neural network, image retrieval algorithm, image recommendation algorithm, implicit support vector machine.

I. INTRODUCTION

With the continuous development of information technology and network platform, various network platforms, including images and other information, will produce a large amount of data. Faced with such a large amount of data, users want to be able to quickly find the relevant image information of interest. Therefore, various network platforms provide image retrieval functions based on this requirement. However, in this kind of search, users need to initiate service requests, and the recommendation system will appear at historic moments to improve the quality of service.

Recommend the most concerned personal information and image data in the image database to users, so that users can obtain the most interested picture data, that is, recommend the most close to the interesting picture to users. In the Internet big data environment, to solve the problems of low image classification accuracy and slow search speed, the research on image classification and search through relevant technologies is now an opportunity and challenge in the image field, which is of great importance in the specific social business value and academic research.

Liang X's team used a novel up-down cultural convolutional neural network (co-cnn) architecture to solve human parsing tasks, which well integrated cross-layer context, global image-level context, semantic edge context, super-pixel in-context and cross-super-pixel neighborhood context into a unified network. Their joint CNN has a significant advantage over other advanced technologies in human parsing [1]. Although the joint CNN method proposed by them has significant advantages, the comprehensiveness of the algorithm should be considered. Wang T's team proposed a composite clustering method based on feature points. This method includes a combination of the feature point

The associate editor coordinating the review of this manuscript and approving it for publication was Zhihan Lv.

aggregation algorithm and the algorithm based on the local color histogram construction strategy. Compared with the existing local color histogram retrieval methods based on feature points, this method can effectively solve the problem of over-dependence on the location information of feature points and the center of feature points. Their research improved the performance of traditional image retrieval methods and enhanced the feasibility of the algorithm [2]. Although their research has improved the performance of traditional methods, the speed of retrieval needs to be further improved. Yu L's team is working on a product image recommendation system based on image content. Their study consisted of three phases. First, the image is preprocessed by removing the background. Secondly, a weighted representation model is proposed to represent images. Finally, the feature combination search scheme of the recommended image is presented. They found that the algorithm they proposed had high accuracy in rapidly recommending similar product images [3]. Although the algorithm proposed by them has high accuracy, it needs to consider the size of the database.

In this paper, image classification algorithm is firstly studied. LReLU - Softplus activation function is formed by combining LReLU function and Softplus function, and CNN is improved. Then an image retrieval model based on local sensitive hashing algorithm is proposed. Then, an image recommendation model based on implicit support vector machine (SVM) is proposed in this paper. This image recommendation method combines image text information and image content information. The experimental results show that the proposed image recommendation model can meet the practical needs.

II. IMAGE RECOMMENDATION ALGORITHM

A. IMAGE CLASSIFICATION ALGORITHM BASED ON IMPROVED CONVOLUTIONAL NEURAL NETWORK

1) CONVOLUTIONAL NEURAL NETWORK

Convolutional neural network (CNN) has more advantages than other neural networks in the ability to process images. First of all, the network topology of CNN and the input image can better match. Secondly, when processing images, CNN can carry out feature extraction and image classification at the same time. This is difficult for a normal neural network to do. Finally, different from the general neural network, CNN's unique concept of weight sharing can greatly reduce the training parameters of the network in the training process, so the overall structure of the convolutional neural network is more adaptable and simpler [4], [5]. Different from artificial neural network, convolutional neural network learns and extracts image features through multi-layer convolution operation and sub-sampling operation. In addition, CNN conducts sub-sampling operation of sampling processing on the feature map obtained after convolution operation, which not only reduces the image dimension but also retains useful feature information. After multiple convolution and sub-sampling operations on the input image, the image features are finally output in a fully connected manner. Then, the error between the output result and the sample label is calculated, and the network weight is updated several times with the back-propagation algorithm [6]. The structure of the convolutional neural network is shown below.

a: INPUT LAYER

The data input layer usually carries out the input and simple preprocessing of the original image, and the preprocessing of the image generally adopts methods such as de-averaging, normalization, PCA and whitening [7]. PCA is a common data dimension reduction operation. Whitening refers to removing the correlation between the data and normalizing the amplitude of the axis.

b: CONVOLUTIONAL LAYER

The convolutional neural network is mainly composed of a stack of convolutional layers with a specific number of channels. Feature extraction and feature mapping are carried out by convolution checking the convolution operation of incoming data, and weighted sum of multiple feature maps is taken into the activation layer to obtain the feature mapping of nonlinear feature extraction [8]. For the image under the two-dimensional data representation, the input is X and the convolution kernel is W, then the convolution operation is shown in formula (1):

$$S(i,j) = (X * W)(i,j) = \sum_{m}\sum_{n} X(m,n)W(i-m,j-n) \tag{1}$$

c: RELU EXCITATION LAYER

CNN ReLU commonly used excitation function (The rectified linear unit), ReLU calculation is simple and has a faster convergence speed, and easy to gradient, compared with Sigmoid function Tanh, effectively reduce The gradient disappear, will be able to train further network [9]. ReLU modified linear element is used as excitation function to carry out nonlinear mapping on the output results of convolution layer. ReLU activation function is defined as formula (2), whose range is $[0, +\infty)$:

$$f(x) = \max(0, x) \tag{2}$$

d: POOLING LAYER

In general, for the convolution layer with a small sensing field and step size, the feature map obtained after convolution is larger. Then, a pooling layer is put into the continuous convolutional layer. The feature map is dimensionally reduced in the following sampling method on the premise of keeping the depth features unchanged, and the feature data output from the convolutional layer is sampled, dimensionally reduced and the regional feature is compressed. It reduces the number of training parameters, speeds up the calculation speed and avoids the over-fitting phenomenon.

e: FULLY CONNECTED LAYER

Generally, after the last pooling layer, several full connection layers are connected as hidden layers. Each neuron in the fully connected layer is fully connected with the neuron in the previous layer, which can integrate the local information in the convolutional layer or pooling layer, and reduce the loss of feature information in the way of refitting.

f: OUTPUT LAYER

The last fully connected hidden layer is connected to the output layer for classification or regression tasks, which is used to output the final results of the neural network.

2) TRAINING OF CONVOLUTIONAL NEURAL NETWORK

Convolutional neural network uses model training to learn the mapping relationship between input and output from a large number of inputs and outputs. The key to learning is to set the weights between layers of the neural network in the training process, and update the mapping relationship by updating the weights [10]. In the process of forward propagation, the partial derivative of the loss function with respect to each weight, the gradient, is mainly calculated by chain derivative, and then the weight of the neuron is updated forward by means of backward propagation according to the gradient descent formula, so as to optimize the network [11].

a: FORWARD PROPAGATION

Forward propagation refers to a sample that is passed into the neural network. If the weight of each layer of the network is W,and the mapping function is F, the output result will be matrix multiplication of weight in the form of sum after passing through the layers, and the data will arrive at the output layer. The output formula is shown in formula (3):

$$O_p = F_n(\ldots(F_2(F_1(X_p W^{(1)})W^{(2)})\ldots)W^{(n)}) \tag{3}$$

b: BACK PROPAGATION

Back propagation means that the error loss between the output value obtained by the neural network and the actual output value is propagated back, and the weight coefficient obtained in the forward propagation is adjusted to minimize the final cost function. For sample n, the calculation formula of back-propagation error is:

$$E^n = \frac{1}{2}\sum\nolimits_{k=1}^{C}(t_k^n - y_k^n)^2 = \frac{1}{2}\left\|t^n - y^n\right\|_2^2 \tag{4}$$

where, C is the total number of categories of the sample set, t_k^n is the kth classification standard in the NTH sample data, and y_k^n is the kth output data corresponding to the nth sample value.

The weight update process of the back-propagation algorithm is as follows:

For a given sample set D = {(x,t)}, initialize the network structure $d \times n_H \times c$. The weight system CO was assigned a random initial value, the learning rate was η, and the threshold value was θ.

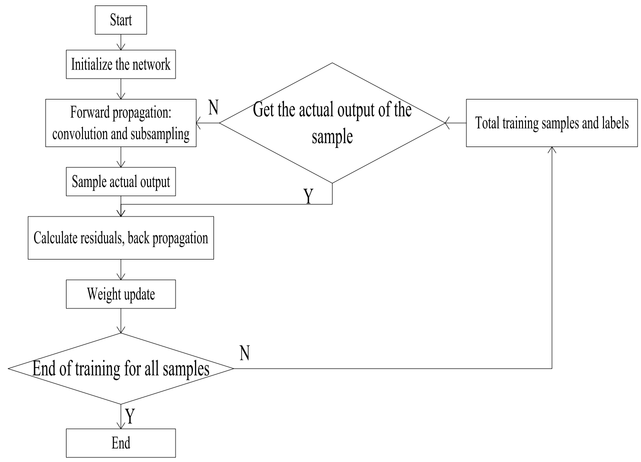

FIGURE 1. CNN training process.

For the kth sample (x,t) in the sample set, the neuron weight was updated.

Iterate through the loop several times until the expected average misalignment is achieved for the sample set. The calculation of its error is shown in formula (5):

$$J = J_x(\omega) = \frac{1}{2}\sum\nolimits_{k=1}^{c}(t_k - z_k)^2 \tag{5}$$

The training process flow chart of CNN is shown in Figure 1.

3) IMPROVED CONVOLUTIONAL NEURAL NETWORK

Although LReLU function can solve the problems of numerical deviation and neuron death of ReLU function, it lacks smoothness. Although Softplus function has smoothness, its output offset phenomenon will affect the convergence of the network. Based on the characteristics of LReLU function and Softplus function, this paper proposes an improved activation function combining LReLU and Softplus functions, which we call lrelu-softplus function. The expression is shown in formula (6), in which, lep is a constant close to 0.

$$f(x) = \begin{cases} ax, & x \le 0 \\ \ln(e^x + 1) - \ln 2, & x > 0 \end{cases} \tag{6}$$

4) IMPROVE THE ACTIVATION FUNCTION OF CNN

In CNN, each layer of neurons uses the activation function to define the output of the neuron, and the output layer is followed by the image classification. Therefore, the Sigmoid function is generally used in the output layer and the full connection layer. Based on this, this chapter discusses the model network performance of the convolutional layer and the sub-sampling layer using the improved lrelu-softplus function, while the full connection layer and the output layer using the Sigmoid function. The CNN structure of the improved activation function is shown in Figure 2.

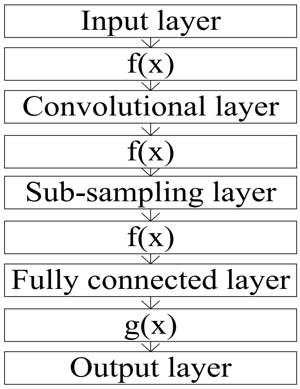

FIGURE 2. CNN structure with improved activation function.

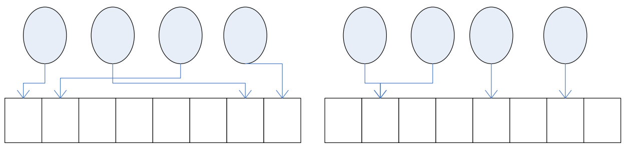

FIGURE 3. Comparison of two hash methods.

B. IMAGE RETRIEVAL ALGORITHM

1) HASH LAYER

For the construction of hash layer after convolutional neural network, its core idea is Locality Sensitive Hashing (LSH), also known as position Sensitive hash, which is mainly used to solve the problem of fast retrieval of high-dimensional mass data [12]. The basic idea is to design a special hash function for high-dimensional mass data. For two data points with high similarity, there is a high probability of mapping to the same hash code. If the similarity between the two data points is low, the probability of mapping to the same hash code is low [13]. When carrying out high dimensional data retrieval, the retrieval results are obtained by calculating the hamming distance between the data points and the points in the retrieval database and sorting the size. The idea of locally sensitive hash is to store the hash codes of the points close to each other in the near hash table, and the farther away the features are from each other in the different hash tables. This is different from the conventional hash algorithm, whose idea is to store them as widely as possible to avoid the conflict between hash codes [14]. The method of local sensitive hash is adopted. In the retrieval, the local sensitive hash only needs to hash the features of the image to be retrieved, so that the images indexed by the features stored in the linked list of the corresponding hash table are the most similar to the image to be retrieved, that is, the similar results of the final retrieval are in this image set. Comparison of two hash methods is shown in Figure 3.

For the hash function under the local sensitive hash algorithm, the spectral hash method proposes that the whole data set should be encoded with as few bits as possible, that is, the generation of compact binary code is required, and the independence between different hash functions is guaranteed. The reason for considering the local sensitivity algorithm is that the similarity measure of the binary code generated by this algorithm in the vector form in hamming space is fast in calculation and can achieve excellent retrieval speed performance. Using convolution neural network learning image characteristics and hash code function at the same time, to adjust classification model, the model after introducing hash layer, with clothing image categories tags at the same time learning image characteristics and hash function, classification error back propagation training model of hash function, make clothing image characteristics of generated binary hash code for the classification and retrieval of minimum error. In the process of retrieval, the binary hash code generated by the feature extraction of the image to be retrieved under the hash layer mapping is used to effectively retrieve the large-scale image data in the low-dimensional hamming space.

2) HAMMING DISTANCE

The similarity measurement of clothing images is transformed into the corresponding binary hash encoding measurement by hash method. In the matching process of retrieval, the hamming distance is used to measure the similarity between clothing images [15]. Hamming distance is used to measure the similarity of binary vectors in hamming space. For two binary feature vectors of length n, u and v, hamming distance is defined as the number of two feature vectors in the same position but with different binary codes. The calculation formula of hamming distance is as follows:

$$D(u, v) = \sum\nolimits_{i=1}^{n} xor(u_i \neq v_i) \tag{7}$$

where, xor is the xor operation,$u_i$and v_i is the ith dimension characteristic vector value of u and v.

3) EUROPEAN DISTANCE IMPROVEMENT

The local sensitive hash algorithm is used to map the high-dimensional feature data of clothing images into binary hash codes, which solves the problem of dimension disaster to some extent [16], [17]. However, when the similarity measure is used to obtain the result of top-k, the comparison of hamming distance sometimes results in an unsatisfactory situation. In order to improve the accuracy of retrieval, the similarity measurement of Euclidean distance in hash bucket is constructed with the guarantee of retrieval speed. The calculation formula of Euclidean distance is as follows:

$$L(A, B) = \sqrt{\sum\nolimits_{i=1}^{n} (a_i - b_i)^2} \tag{8}$$

For the similarity measurement results obtained in hamming space, top-k, the Euclidean distance calculation between the image to be retrieved and the image feature vector of the top-k image set, and the ranking results of the Euclidean distance measurement were used as the final retrieval results.

C. IMAGE RECOMMENDATION ALGORITHM BASED ON IMPLICIT SUPPORT VECTOR MACHINE

1) SELECTION AND EXTRACTION OF IMAGE FEATURES

a: IMAGE COLOR FEATURE EXTRACTION

Color feature is the most widely used visual feature in image recommendation and retrieval. Compared with other visual features, color is highly correlated with the scene objects contained in the image and has high robustness [18]. Therefore, color is taken as one of the extracted features in the recommendation algorithm of content personalization based on image. The main steps of color feature extraction are as follows: first, select the appropriate color space to describe the color features; Then, the color feature is converted into the form of feature vector by the quantitative method.

RGB color space is the most commonly used color space in image processing, and the HSV color space, HIS color space and LAB color space we often use in image processing are all converted from RGB color space [19]. However, RGB has some shortcomings: it is not intuitive enough, and there is a big gap between the difference between the two colors calculated by euclide distance and the difference between the two colors actually observed by people. HSV color space is adopted in this paper, and the color features of images are extracted by means of global histogram segmentation.

b: PHOG FEATURE EXTRACTION

PHOG feature extraction is a feature describing the spatial shape, which can not only represent the overall shape of the image, but also the local shape of the image and the spatial relationship between the whole and the part [20], [21]. The core idea of PHOG feature extraction is to first layer the whole image, then extract HOG features in each layer, and finally cascade the layered HOG features to obtain PHOG features.

2) IMAGE RECOMMENDATION ALGORITHM BASED ON IMPLICIT SUPPORT VECTOR MACHINE

a: IMPLICIT SUPPORT VECTOR MACHINE

Support Vector Machine (SVM) is a supervised Machine learning method based on statistical learning theory, which has become the most widely used at present. The main ideas of SVM are as follows: first, in the case of linear separability, SVM can be analyzed; In the case of linear inseparability, relaxation variables are added for analysis, and the nonlinear mapping algorithm is used to transform the linear inseparability samples in the low-dimensional input space into the linearly separable samples in the high-dimensional feature space, and then the linear algorithm is used in the high-dimensional feature space for the linear analysis of the samples [22]. Secondly, based on the theory of structural risk minimization,SVM constructs the optimal segmentation hyperplane (classification surface) in the sample space or feature space, so that the distance between the classification surface and non-homogeneous samples is the largest, thus enabling the learner to obtain the global optimal solution [23].

Implicit support vector machine (LSVM) is generated by adding some reference information on the basis of linear SVM, so LSVM will be constrained by these additional parameters in the process of model training [24], [25]. Therefore, in the process of model training, not only the conventional image features should be considered, but also these additional parameter information should be considered. In order for LSVM to find the optimal classification line, these parameters should be optimized by operations outside the classifier. After the optimization, training samples should be sent to SVM for iterative optimization.

b: IMAGE RECOMMENDATION ALGORITHM BASED ON IMPLICIT SUPPORT VECTOR MACHINE

The image I of the user in the training set is defined as an array (x, a_u, o). Among them, $x = [x_p; x_c; x_g]$ is the multi-feature fusion vector of the training samples, x_p is the PHOG feature vector of the image, x_c is the color feature vector of the image, and x_g is the Gist feature vector of the image. O is the image content classification result of the training sample, through multiple instance learning algorithm of the image tag to the regional level, and using the segmentation technology, the label and corresponding to the image of interest in the area, the corresponding label and image area is defined as the vector $a_u = [a_1^u, a_2^u, \ldots, a_{k_u}^u]^T$, the k_u is the marked area number.

The main goal of the lsvm-based image personalization semantic analysis recommendation algorithm is to learn a model that can be used to recommend the image that the user might be most interested in. In the experiment, for N training sample sets $\{x^{(n)}, a_u^{(n)}, o^{(n)}\}_{n=1}^N$, we constructed a scoring function to score and sort the samples in the test set and recommend the sample image with the highest score as the target. The scoring function is:

$$f_w : \chi \times \theta \to R \tag{9}$$

In the above formula, X is a sample image in the training set,w is a parameter of the scoring function f_w, and X represents the user's historical image space. During the test, function f_w is used to select the image X^+ that users are most interested in from all areas of the panoramic image χ^t classified by a given scene. X^+ is defined as:

$$X^+ = \arg\max_{X \in X^t} f_w(x, o) \tag{10}$$

Then the recommended image of all users in the test set can be expressed as:

$$X_s^+ = arc\max_{X \in X^t} f_w([x_a, x_s], o) \tag{11}$$

where, X_s^+ is to select the image region that users may be most interested in from the panorama, x_a is the user's historical image, and x_s is the panorama image. In this paper, based

TABLE 1. Experimental environment.

Type	Name	Version or size
Operating system	Linux	Manjaro
Development tool	Pycharm	4.5.4
Development language	Python	3.7
Development framework	PyTorch	1.4

on the implicit support vector machine, we assume that the scoring function $f_w(x, o)$ is defined as:

$$f_w(x, o) = \max_{a_u} w^T \varphi(x, a_u, o) \tag{12}$$

φ is a fusion feature vector of image x, a_u is an image label set,o is an image content classification result set, and w is a model parameter of LSVM. In this model, the defined recommendation function can be expressed as:

$$w^T \phi(x, a_u, o) = w_o^T \delta(x, o) + \sum_{j \in A^m} w_{a_j}^T \omega(x,a_j) \tag{13}$$

In this type, w a parameter vector is all parameters combination factors, at the same time considering the dependencies between the visual features and scene, the first w represents the forecast panoramic scene visual characteristics, the second w describes user history scene image and the panorama feature value of similar degree, $\delta(x, o)$ delegates from user history images to extract the feature vector x specific mapping, and the result of mapping of label o depends on the scene. The potential function $\omega(x, a_j)$ represents the trust obtained by standard multi-level linear SVM.

III. EXPERIMENTS OF IMAGE RECOMMENDATION ALGORITHM

A. DATA COLLECTION

In order to obtain a more real experimental effect, this section of experimental data set adopts 10 categories of women's dress images selected from the e-commerce platform. Referring to the fine category setting of women's wear on these e-commerce platforms, the data set selects the top 10 categories with large amount of data for classification setting, which can also be extended in the later stage. There are 100 images in each category, almost all of which are based on commodities. There no problems such as different perspectives, insufficient light, blurring and shielding, etc. There are a total of 1,000 images. The average size of each image is 200*200, and the format is JPEG. Try to select the image with big difference between each type of sample.

The Mnist handwritten numeral library contains 70,000 images. In the experiment on the Mnist library, 60,000 images were selected as samples, and the remaining 10,000 images were used for testing. The preprocessing only needs to normalize the image to [0,1].

B. EXPERIMENTAL ENVIRONMENT

Processor: IntelCorei5-9400F@6 $\times$ 4.1GHz; Graphics: GeForceGTX1660Ti; Memory: 16 g; Operating system :Manjaro Linux environment; Development environment: based on Python3.7 and PyTorch1.4. The specific experimental environment is shown in Table 1.

C. EXPERIMENTAL DESIGN

In the learning process of image recommendation model based on implicit support vector machine, the extraction of feature vector is divided into two parts: feature extraction of user history image and feature extraction of panorama image.

The image feature extraction of panorama is divided into two steps: first, the panorama is divided into several regions, and then the features of each region are extracted respectively. Size the user's area between a square area of 20*20 pixels and an area of 100*100 pixels. To ensure that the window area can contain all the contents of the panorama, we set the area block as a sliding area block of N*N pixels. We slide from the top left corner of the panorama, from left to right, from top to bottom, and move the distance of M(M $\leq$ N) pixels each time. With a sliding window, you don't miss out areas of the image in the panorama that weren't taken into account. Area for 50*50 pixels in sliding window, for example, with 5 pixels as the step length, panorama of the upper left corner, from left to right, from top to down, the image based on LSVM recommended method, can be based on user history image personalized analysis, thus from the panorama of the same scene classification results to find out the recommended area, users are most interested in.

The manual recommendation method uses 20 people who are interested in taking photos to directly observe and recommend the panorama area. The analysis of the manual recommendation area shows that the recommended area of most people is similar, while the recommended area of a small number of people is very different.

IV. EXPERIMENTAL RESULTS OF IMAGE RECOMMENDATION ALGORITHM

A. ANALYSIS OF IMAGE CLASSIFICATION TEST RESULTS ON DIFFERENT DATA SETS

1) EXPERIMENTAL RESULTS ON THE COMMODITY IMAGE LIBRARY

In order to obtain a more real experimental effect, this section of experimental data set adopts 10 categories of women's dress images selected from the e-commerce platform. The experimental results of the image classification algorithm based on improved CNN proposed in this paper on the commodity image library are shown in Table 2 and Figure 4.

It can be seen from Table 2 and Figure 4, on the commodity image library, the model recognition effect of lrelu-softplus activation function proposed in this paper is better than that of

TABLE 2. Experimental results on the commodity image library.

	10	20	30	40	50	60
ReLU	50	52	78	80	84	86
Softplus	51	52	80	82	85	87
Tanh	52	53	81	84	87	89
Sigmoid	53	54	82	85	88	90
LReLU	54	56	83	86	90	92
LReLU-Softplus	55	60	89	95	96	97

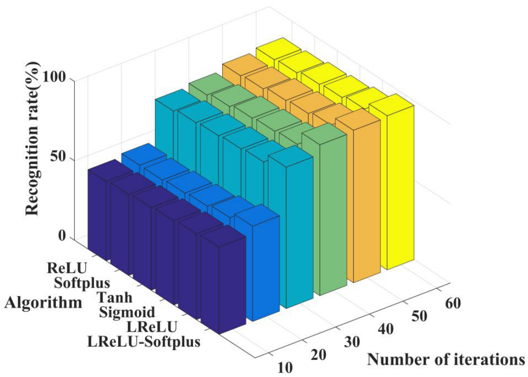

FIGURE 4. Experimental results on the commodity image library.

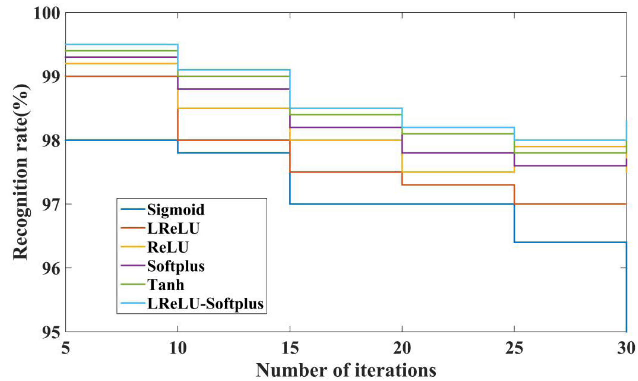

FIGURE 5. Experimental results on the handwritten digit library.

other activation functions. With the increase of the number of iterations, the recognition rate of each activation function has little difference on the whole. When the number of iterations reaches 40, the recognition rate of each function tends to increase gradually. Sigmoid and Tanh are generally lower than ReLU, LReLU and Softplus in terms of recognition performance, which also indicates that the activation function of similar biological nerve can indeed improve the classification performance of the model in the commodity image library. Using the improved lrelu-softplus function is better than LReLU and SoRplus, because the improved activation function makes up for the disadvantages of the original two methods and has the advantages of both.

2) EXPERIMENTAL RESULTS ON HANDWRITTEN NUMBER LIBRARY

The training data set of Mnist handwritten number library is composed of the handwritten Numbers of different people, so there are some differences between the pictures. The experimental results of the image classification algorithm based on improved CNN presented in this paper on the handwritten number library are shown in Figure 5.

As can be seen from Figure 5, on Mnist handwritten numeral library, the CNN recognition effect of LReLU-Softplus activation function proposed in this paper is better than that of other activation functions. With the increase of the number of iterations, the recognition rate of each activation function is obviously different, especially when the number of iterations increases to 25, the effect is more obvious. It is obvious that Sigmoid and Tanh, compared with ReLU, LReLU and SoRplus in terms of recognition performance, further prove the benefits of approximate biological activation function. Again, using the improved lrelu-softplus function works better than LReLU and Softplus. In addition, the recognition rate of the experimental results is higher than that of other image libraries, because there are more training samples in Mnist library, so the recognition rate is higher in the same iteration times.

B. IMAGE RETRIEVAL AND RECOMMENDATION ALGORITHM TESTING

1) ANALYSIS OF TEST RESULTS OF IMAGE RETRIEVAL ALGORITHM

For image retrieval, it is necessary to measure the similarity of extracted image features and return the ranking results. Currently, the linear matching method of features is difficult to learn and easy to form dimension disaster, resulting in poor retrieval effect. In this paper, the method of hash is considered to add a hash layer after feature extraction of the garment image classification model, input image features into the hash layer, learn the hash function to generate binary hash codes, measure the distance based on the binary hash codes of the hash function, and get the sorting result as the result of retrieval. For a retrieval, the binary hash code under the feature hash function of the extracted garment image was measured for similarity, and the sequence was performed in combination with hamming distance and Euclidean distance. The results obtained were passed into the user retrieval module as the return results of the retrieval module. The test results of image retrieval algorithm are shown in Figure 6.

TABLE 3. The coincidence degree of the two recommended images.

Region size	20	30	40	50	60	70	80	90	100
PHOG+Color	33	45	67	72	86	83	88	79	64
PHOG+Gist	21	32	39	50	66	41	38	24	25
Color+Gist	21	37	58	79	82	75	80	59	33

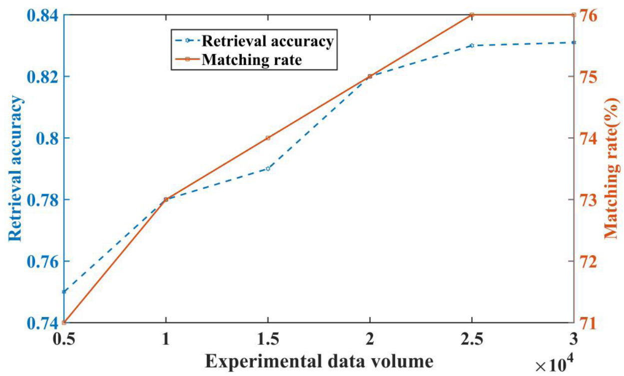

FIGURE 6. Test results of image retrieval algorithm.

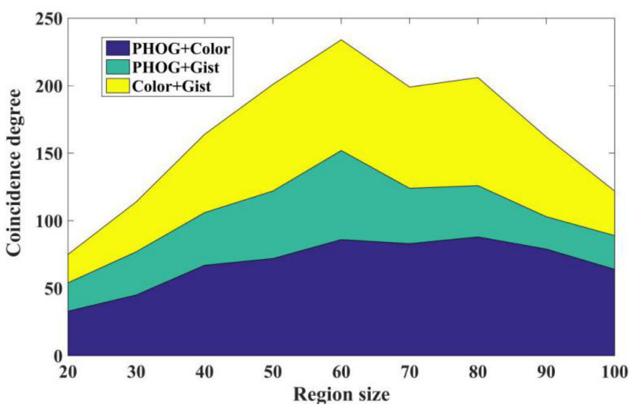

FIGURE 7. The coincidence degree of the two recommended images.

As can be seen from Figure 6, with the expansion of the data set, the retrieval accuracy synchronization is improved due to the increase of samples. When the data set is greater than 20,000, the retrieval accuracy is greater than 0.8. However, the model tends to be stable gradually, which ensures the good stability of the model for the super-large scale similar garment image set. For the retrieval sorting effect, the matching percentage of retrieval is greater than 70% for different data volumes, and the average matching percentage reaches about 74%. It shows that this algorithm has a good retrieval sorting effect.

2) IMAGE RECOMMENDATION ALGORITHM TEST

In this paper, the degree of overlap between the image recommended by the personalized semantic analysis method and the image recommended by the artificial recommendation method is used to test the accuracy of the proposed algorithm. The experiment found that the matching degree of the two methods reached 88%. At the same time, it is also found that different size of panorama sliding window has a great impact on the recommended results. The coincidence degree between the recommended images in sliding windows of different sizes and those recommended by manual recommendation method is shown in Table 3 and Figure 7.

As can be seen from Table 3 and Figure 7, when the size of the sliding window is set to 60*60 pixels to 80*80 pixels, the overlap degree of the two methods is the largest, exceeding 85%, indicating that the size of the sliding window is directly related to the image overlap rate of the two recommended methods. The extraction of different image features also has an impact on the coincidence rate. It can be observed that the fusion degree of PHOG and Color features is larger than that of other two feature fusion images.

V. CONCLUSION

Based on the characteristics of LReLU function and Softplus function, a convolutional neural network with improved activation function is proposed in this paper, which is called LReLU-Softplus function. The improved lrelu-softplus activation function uses the LReLU function in the part less than 0, whose value is close to 0 but not completely 0, so the output offset problem of the Softplus function can be solved well. The parts greater than 0 use Softplus functions. Has the smoothness, compensates lacks the smoothness LReLU function the shortcoming. Experimental results show that the lrelusoftplus function proposed in this paper is better than other activation functions.

The proposed image retrieval model, compared with the traditional content-based image retrieval technology, using improved local sensitive hash algorithm, to deep learning approach to study hash function to generate a binary hash code, binary hash code under the hamming space similarity measure to speed up the speed of image retrieval, in theory, should have good retrieval effect.

This paper proposes and implements an image analysis recommendation algorithm based on implicit support vector machine. Different from existing image recommendation methods, this personalized image recommendation method combines image text information and image content information. The experimental results show that the proposed image recommendation method based on implicit support vector machine can meet the personalized needs of different users.

ACKNOWLEDGMENT

The financial support is gratefully acknowledged.

REFERENCES

[1] X. Liang, C. Xu, X. Shen, J. Yang, J. Tang, L. Lin, and S. Yan, "Human parsing with contextualized convolutional neural network," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 1, pp. 115–127, Jan. 2017.

[2] T. Wang and W. Wang, "Research on new multi-feature large-scale image retrieval algorithm based on semantic parsing and modified kernel clustering method," *Int. J. Secur. Its Appl.*, vol. 10, no. 1, pp. 139–154, Jan. 2016.

[3] L. Yu, F. Han, S. Huang, and Y. Luo, "A content-based goods image recommendation system," *Multimedia Tools Appl.*, vol. 77, no. 4, pp. 4155–4169, Feb. 2018.

[4] P. Moeskops, M. A. Viergever, A. M. Mendrik, L. S. de Vries, M. J. N. L. Benders, and I. Isgum, "Automatic segmentation of MR brain images with a convolutional neural network," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1252–1261, May 2016.

[5] B. Yang, J. Zhong, Y. Li, and Z. Chen, "Multi-focus image fusion and super-resolution with convolutional neural network," *Int. J. Wavelets, Multiresolution Inf. Process.*, vol. 15, no. 4, pp. 136–146, 2017.

[6] F. Lu, F. Wu, P. Hu, Z. Peng, and D. Kong, "Automatic 3D liver location and segmentation via convolutional neural network and graph cut," *Int. J. Comput. Assist. Radiol. Surgery*, vol. 12, no. 2, pp. 171–182, Feb. 2017.

[7] S. S. Mohseni Salehi, D. Erdogmus, and A. Gholipour, "Auto-context convolutional neural network (auto-net) for brain extraction in magnetic resonance imaging," *IEEE Trans. Med. Imag.*, vol. 36, no. 11, pp. 2319–2330, Nov. 2017.

[8] Z. Zhang, H. Wang, F. Xu, and Y.-Q. Jin, "Complex-valued convolutional neural network and its application in polarimetric SAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 7177–7188, Dec. 2017.

[9] I. Sevo and A. Avramovic, "Convolutional neural network based automatic object detection on aerial images," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 5, pp. 740–744, May 2016.

[10] C. Yuan, X. Li, Q. Wu, J. Li, and X. Sun, "Fingerprint liveness detection from different fingerprint materials using convolutional neural network and principal component analysis," *Comput., Mater. Continua*, vol. 53, no. 4, pp. 357–371, 2017.

[11] Q. Wang and R. Zhang, "Double JPEG compression forensics based on a convolutional neural network," *EURASIP J. Inf. Secur.*, vol. 2016, no. 1, pp. 1–12, Dec. 2016.

[12] Y. Liu, Y. Peng, K. Lim, and N. Ling, "A novel image retrieval algorithm based on transfer learning and fusion features," *World Wide Web*, vol. 22, no. 3, pp. 1313–1324, May 2019.

[13] R. Zeng, Y. Wang, and W. Wang, "A co-occurrence region based Bayesian network stepwise remote sensing image retrieval algorithm," *Earth Sci. Res. J.*, vol. 22, no. 1, pp. 29–35, Jan. 2018.

[14] R. Jegadeeswari and V. Khanna, "An integrated approach for content based image retrieval system using wavelet and fuzzy C-means algorithm," *Int. J. Control Theory Appl.*, vol. 9, no. 28, pp. 133–140, 2016.

[15] Q. Zhang, Q. H. Lin, and X. Kang, "Research on image retrieval based on kernel density estimation and fractal coding algorithm," *Jiliang Xuebao/Acta Metrologica Sinica*, vol. 38, no. 3, pp. 284–287, 2017.

[16] G. H. Wei, S. Qi, W. Qian, and K. Zhang, "Image retrieval algorithm of pulmonary nodule based on similarity measurement," *Dongbei Daxue Xuebao/J. Northeastern Univ.*, vol. 39, no. 9, pp. 1226–1231, 2018.

[17] B. S. Phadikar, A. Phadikar, and G. K. Maity, "Content-based image retrieval in DCT compressed domain with MPEG-7 edge descriptor and genetic algorithm," *Pattern Anal. Appl.*, vol. 21, no. 2, pp. 469–489, May 2018.

[18] D. Sejal, V. Rashmi, K. R. Venugopal, S. S. Iyengar, and L. M. Patnaik, "Image recommendation based on keyword relevance using absorbing Markov chain and image features," *Int. J. Multimedia Inf. Retr.*, vol. 5, no. 3, pp. 185–199, Sep. 2016.

[19] J. Zhang, Y. Yang, Q. Tian, L. Zhuo, and X. Liu, "Personalized social image recommendation method based on User-Image-Tag model," *IEEE Trans. Multimedia*, vol. 19, no. 11, pp. 2439–2449, Nov. 2017.

[20] J.-H. Chung and Y.-J. Park, "The effect of service quality and service value on brand image and recommendation intention of business hotel," *Korean J. Hospitality Tourism*, vol. 26, no. 5, pp. 53–69, Jul. 2017.

[21] R. A. Wahurwagh and P. M. Chouragade, "Popular place prediction and image recommendation using hierarchical multi-clue modeling for tourist," *Int. J. Comput. Sci. Eng.*, vol. 7, no. 4, pp. 969–972, Apr. 2019.

[22] Y.-F. Lu, H. Qiao, Y. Li, and L.-H. Jia, "Image recommendation based on a novel biologically inspired hierarchical model," *Multimedia Tools Appl.*, vol. 77, no. 4, pp. 4323–4337, Feb. 2018.

[23] T. Khan, "Smart microwave oven with image classification and temperature recommendation algorithm," *Int. J. Electr. Comput. Eng. (IJECE)*, vol. 8, no. 6, p. 4239, Dec. 2018.

[24] G. Haifeng, S. Shoubao, and S. Zhoubao, "Image tag recommendation based on friendships," *Multimedia Tools Appl.*, vol. 76, no. 13, pp. 1–17, 2017.

[25] Y. Gu, C. Wang, J. Ma, R. J. Nemiroff, D. L. Kao, and D. Parra, "Visualization and recommendation of large image collections toward effective sensemaking," *Inf. Visualizat.*, vol. 16, no. 1, pp. 21–47, Jan. 2017.

PEI YIN was born in Chongqing, China, in 1987. She received the Ph.D. degree from Tongji University, China. She is currently working with the Business School, University of Shanghai for Science and Technology. Her research interests include machine learning, deep learning, and recommendation systems.

LIANG ZHANG was born in Hebi, Henan, China, in 1988. He received the Ph.D. degree in engineering mechanics from the Institute of Mechanics, Chinese Academy of Sciences, Beijing, China, in 2016. He is currently a Postdoctoral Fellow with the School of Aerospace Engineering, Tsinghua University. His current research interests include structural dynamics, noise and vibration control, smart material, and structure.