Construction of Image Retrieval Module for Ethnic Art Design Products Based on DF-CNN

Yaru He
Chengdu Vocational and Technical College, Chengdu, 610041, China

***Abstract*—With the increasing interest of consumers in ethnic art, more design products with ethnic art characteristics are being displayed. In order to help users easily retrieve related art products, an image retrieval model that can effectively extract data is proposed. The research method strengthens the depth of data mining through weighted methods, main characteristics and local features in images based on the multi-window combination, and uses the deep forest algorithm to expand the decision path and select information gain nodes. By adjusting the weights of convolutional neural networks, the retrieval ability of the model is enhanced. The gradient problem in the propagation process is optimized using residual modules, and the prominent features of the features are strengthened using a bar attention mechanism to optimize the retrieval ability. The results indicated that the loss function of the research model converged within 20 iterations, and the matching degree of the retrieved images in the testing set reached 91.28% after iterative training. The AUC of the research model was 0.876, indicating that the model had a good performance in image retrieval and classification. The retrieval accuracy of the research model was higher than other methods for image data of different specifications. This indicates that the research model has universality for multi-scale image retrieval, which can provide theoretical support for the development of ethnic art design products.**

***Keywords*—Image retrieval; main characteristics; local features; deep forest; convolutional neural network; bar attention mechanism; residual module**

I. Introduction

Ethnic art contains the cultural heritage and artistic ideas of a nation, and products created based on ethnic art have distinct cultural characteristics [1]. More art museums are using ethnic art design as a featured theme to share related types of exhibits. However, the patterns of ethnic characteristic exhibits are usually formed by simple line and pattern combinations, which do not have obvious characteristics of things, making retrieval difficult [2-3]. In order to retrieve images more effectively, various image retrieval method shave been developed. Early image retrieval relied on textual keywords to express image features, using the main parameters of the image as a reference standard, and obtaining retrieval results through data comparison and visual combination. However, method comparison requires manual labeling, resulting in low retrieval efficiency [4]. With the advancement of retrieval technology, content-based image retrieval methods have been proposed. The image search method can avoid manual errors by comparing image attributes with image features to obtain retrieval results. However, the recognition effect on local features is unstable and easily affected by noise [5]. Therefore, D. Lowe proposed the Scale-Invariant Feature Transform (SIFT) feature method and the Local Binary Pattern (LBP) feature method. The local features of the image are enhanced to deepen the model's memory of feature points. Moreover, it has good recognizability for behaviors such as image rotation and scaling, but the method places too much emphasis on local features, resulting in a lack of global perspective [6]. With the development of intelligent technology, Convolutional Neural Networks (CNN) have been applied in image retrieval. The self-learning ability based on CNN can accurately extract image features and enhance the adaptability of network retrieval through dataset training. However, excessive feature vectors may lead to long training time and slow retrieval efficiency of the network. The research aims to design a retrieval module that ensures the accuracy of image retrieval while improving the speed of image retrieval. To optimize the user's search experience and match the search needs of art galleries or museums. Therefore, the DF-CNN algorithm is innovatively proposed for image retrieval in the field of ethnic art products. It strengthens local features through a bar attention mechanism, combines local features with subject features using an overlap pooling method, and optimizes the retrieval speed of the model through parallel multi-channel convolution. A new ethnic art design product retrieval model is constructed to provide certain technical support for the dissemination of ethnic art.

II. Related Work

With the development of Internet technology, the number of image data in the network is gradually rich, but it is difficult to accurately retrieve the required image in a rich database. Therefore, researchers have conducted research on image retrieval methods. Suganyadevi S et al. proposed a research method based on deep learning for medical image retrieval. The method established a learning system by exploring data information in the medical field, and reduced the dimensionality of functions using deep conventional and extreme learning methods to complete the retrieval and analysis of medical images. The results showed that the proposed method was accurate for retrieving medical image data [7]. Kelishadrokhi M K et al. proposed a research method based on novel local texture descriptors and color features for image retrieval problems in computers. The texture and color information database was used to train the image recognition ability. Effective features were extracted using the local neighborhood difference method. The results indicated that the proposed method had a good corresponding effect on computer content image retrieval [8]. Keisham Net al. proposed a research method based on depth search and rescue algorithm for image retrieval on the Internet. The method used advanced visual feature extraction techniques to enhance the feature points of the image, and optimized the feature clustering of the image through feature fusion and filtering. The results indicated that the proposed method significantly improved the efficiency of image retrieval [9]. Wang W et al. proposed a method based on sparse representation and feature fusion for image retrieval in databases. The method used a generalized search tree to retrieve similar scenes in the image and enhanced local features of the image through sparse coding. The results indicated that the proposed method improved the accuracy of database image retrieval [10]. Ning C et al. proposed a method based on deep metric learning for image retrieval of clothing. The method called similar features from the database through pre-set regional scenarios, and optimized the recognition of clothing patterns based on feature similarity comparison and feature point analysis. The results indicated that the proposed method had a fast retrieval speed for clothing patterns [11].

Zhuang H et al. adopted the Deep Forest (DF) algorithm to address the urban land change. A dimensional space was constructed based on terrain modeling. Advanced features in structural data were mined on the basis of deep learning methods, and the changing state of land was simulated through community comparison. The designed method had a relatively accurate prediction level for land change issues [12]. Hamedianfar et al. proposed a method based on the DF algorithm to address the application issues of remote sensing methods. Multi-scale features were established through shallow machine learning, the network architecture was expanded through a time series strategy and remote sensing method was optimized through multiple training methods. The developed method had a good simulation effect on remote sensing data modeling [13]. Shaaban M A et al. adopted a deep convolutional forest to detect text spam emails. Machine learning was used to perform the basic classification of email types. The dynamic deep integration method automatically adjusted the classification complexity and extracted effective features with the help of classifiers. The results indicated that the proposed method accurately isolated phishing emails [14]. Huang et al. proposed a method based on CNN for fault diagnosis in complex systems. The method used sample feature extraction from multivariate time series for multi-layer transmission, optimized data retrieval through sliding processing of data windows, and combined model training to enhance the diagnostic performance. The results indicated that the proposed method had high prediction accuracy for fault diagnosis [15]. Tayal A proposed a research method based on CNN for the diagnosis of retinal diseases. The method automatically identified disease types through the constructed intelligent learning framework, determined disease categories based on the feature set of medical images, and filtered out interference items through image denoising. The results indicated that the proposed method had an assisting role in the diagnosis of retinal diseases [16].

In summary, Kelishadrokhi et al. enhanced the depth of image feature extraction through local texture and color feature recognition, but it affected the retrieval speed of the image. Ning C et al. used deep metric learning to enhance the speed of retrieval, but due to the single feature extraction method, the matching degree of image retrieval decreased. Moreover, a single CNN algorithm lacks feature type differentiation, which can affect the efficiency of ethnic art image retrieval. However, there is currently limited data on the combined application of CNN and DF algorithms in the field. Therefore, research attempts to achieve synchronous improvement in retrieval performance and retrieval speed through the fusion of the two algorithms.

III. Methods and Materials

A. *Design of Image Retrieval Module for Ethnic Art Products Based on DF*

With the exchange and dissemination of culture, ethnic art patterns have gradually emerged as an artwork in the public eye. The application of ethnic patterns has developed from the initial daily life to ethnic art design [17-19]. Art design products with distinctive features during tourism are often given as souvenirs to friends and family. However, faced with a wide variety of ethnic art patterns, the search process is often dazzling [20-22]. In order to efficiently and accurately retrieve ethnic artworks with diverse patterns, the study introduces the DF for the image retrieval process. The DF algorithm, as an efficient classifier, can accurately extract image features from high-dimensional image data and optimize the model's fitting through its self-learning ability [23-25]. The cascading structure processing process of the DF algorithm is shown in Fig. 1.

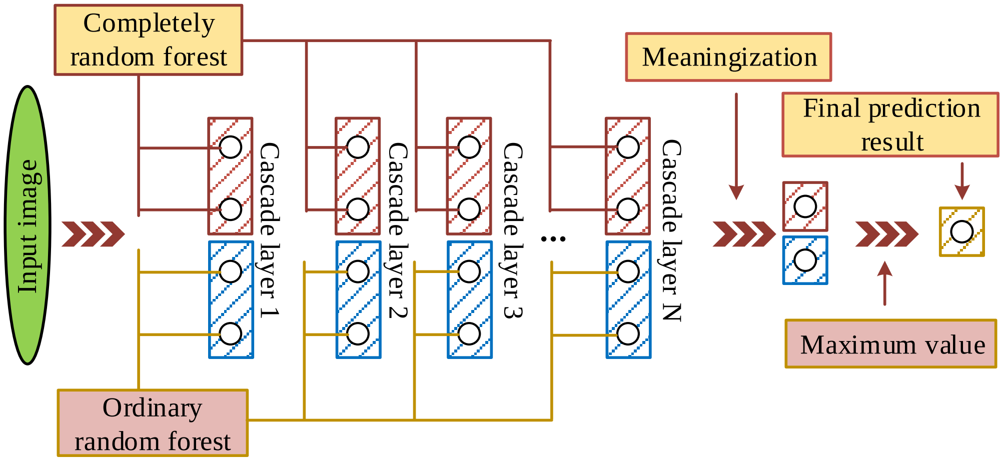


Fig. 1. The cascade structure of deep forests.

In Fig. 1, the DF algorithm processes input data through multiple layers. The process strengthens the data mining depth using the weighted method, and assigns hierarchical weights to the forest based on weight factors. Based on the predicted probability value of the forest, the weight factor's proportion is optimized. Through hierarchical multi-source data synchronization analysis, discrete image features can be effectively detected. To optimize the feature partitioning performance in the DF, the decision tree in the DF algorithm is used for information entropy optimization. The optimization process defines the information entropy of the sample set, as shown in Eq. (1).

$$Ent(N) = -\sum_{i}^{n} \frac{C^i}{N} \log_2 \frac{C^i}{N} \tag{1}$$

In Eq. (1), $Ent(N)$ represents the sample set information entropy. N represents the total sum of the sample set. C^i represents the sample size. The sample ratio is set to $p_i = \frac{C^i}{N}$ during the optimization process. Therefore, the information entropy of the sample set is converted, as shown in Eq. (2).

$$Ent(N) = -\sum_{i}^{n} p_i \log_2 p_i \tag{2}$$

In Eq. (2), p_i represents the sample ratio. Information entropy, as a metric in decision trees, can represent the set purity of a sample set. Based on the difference in information entropy before and after the feature decision dataset, the change in information gain is calculated. The value of information gain is measured by the influence proportion of branch nodes. The gain calculation process is shown in Eq. (3).

$$Gain = (N,a) = Ent(D) - \sum_{v=1}^{V} \frac{|N^v|}{|N|} Ent(N^v) \tag{3}$$

In Eq. (3), $Gain$ represents the information gain of dataset N after being partitioned by feature attribute a. a represents the selected feature attribute. V represents the number of branch nodes that may be formed during the process. N^v represents the sample data when the branch node is V. v represents the range of values for feature attribute a. $|N^v|/|N|$ represents the weight assigned to branch nodes. The feature partitioning of the input image by the decision tree is evaluated by the Gini coefficient, and the evaluation calculation process is expressed as Eq. (4).

$$Gain(D) = \sum_{k=1}^{K} p_k(1-p_k) = 1 - \sum_{k}^{K} p_k^2 \tag{4}$$

In Eq. (4), $Gain(D)$ represents the Gini index. D represents the selected dataset. D represents the probability of decision-making. K signifies the number of branch nodes in the current dataset D. k signifies the value of the current branch node. To extract features from the entire image layer, dynamic sliding windows are used to scan image information in different regions. The complete feature extraction of the image is completed by concatenating the information from overlapping windows. The research process adopts multi-granularity scanning. The principle of the scanning model is shown in Fig. 2.

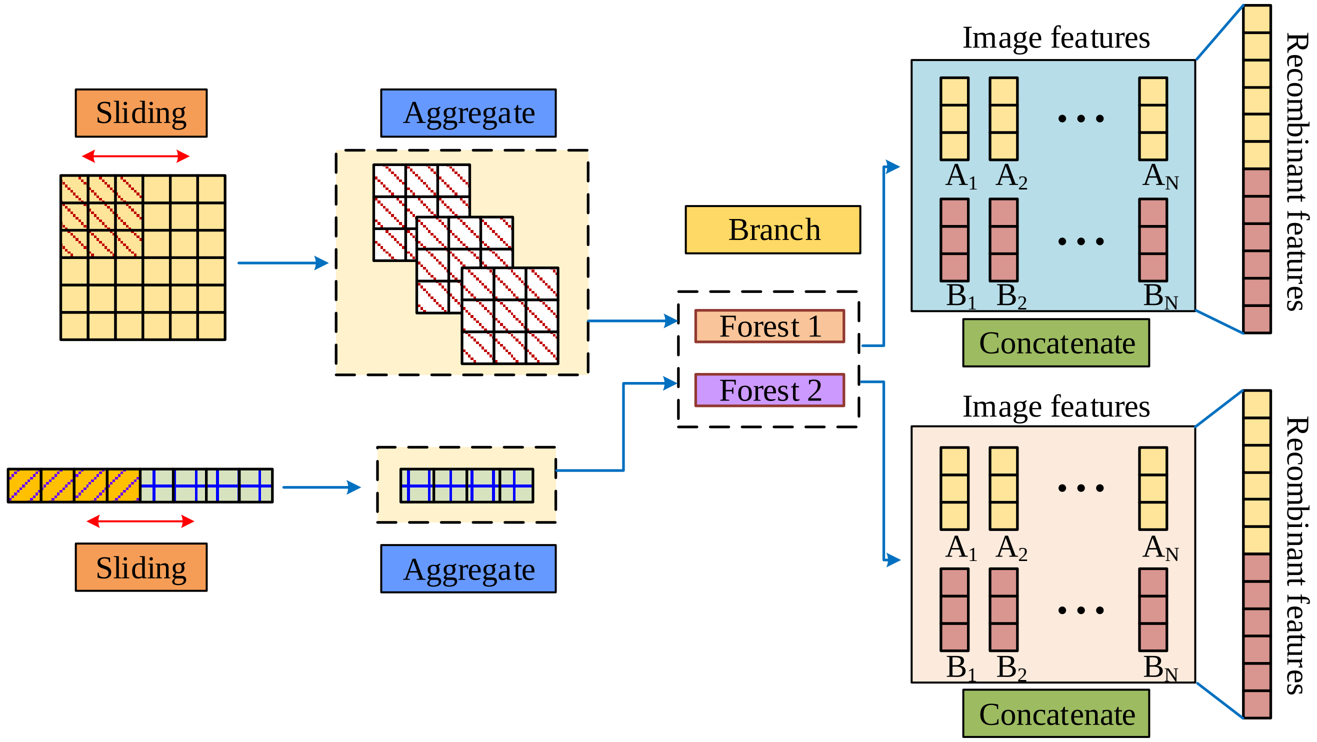


Fig. 2. Multi-granularity scanning process.

In Fig. 2, the multi-granularity scanning process of the DF algorithm uses window frames of different sizes as feature extraction windows. The main and local features in the image are identified through a combination of multiple windows. The feature output extracted by sliding window is used as a probability vector, and the main output and local output are hierarchically placed into a cascaded forest. The image dimension is enhanced through multi-vector scanning and transmission. In order to enhance local data feature processing, both completely random forest and ordinary random forest are used as decision paths, and split nodes with information gain are selected. To address the feature loss caused by dimensional differences, a linear discriminant analysis method is introduced to perform correlation recognition of image features. The mean vector is expressed as Eq. (5) during the calculation process.

$$u_j = \frac{1}{N_j} \sum_{x \in X_j} x\left(j = 0,1 \right) \tag{5}$$

In Eq. (5), u_j represents the mean vector of the classes j in the sample. N_j signifies the number of classes j in the sample. X_i signifies the dimension vector value of the sample. X_j signifies the class sample set of group j. The covariance matrix of sample features is calculated, as shown in Eq. (6).

$$\sum j = \sum_{x \in X_j} (x - u_j)(x - u_j)^T (j = 0,1) \tag{6}$$

In Eq. (6), $\sum j$ represents the covariance matrix of the sample. T represents the transpose of a matrix. In order to calculate the projection point positions of two sets of samples, the divergence matrix is calculated in such a way that similar samples are close and dissimilar samples are far away. The optimization objective is rewrite through high-dimensional to low dimensional vector mapping. The maximum eigen value of the matrix is calculated by the generalized Rayleigh quotient. The DF algorithm generates its own feature vectors at each level when transmitting image features step by step. However, the algorithm's extraction step for the image is based on direct learning of the overall features. Too many levels in the cascaded forest may cause the enhanced features to be covered by ordinary feature vectors, resulting in weak fitting of the model to image samples even after learning. Due to the urgent need for retrieving images of ethnic art products in the scene, the study introduces the prior box method in the scanning stage to make the image extraction process faster. The specific working process of the prior box is shown in Fig. 3.

In Fig. 3, the prior box sets the window size that matches the extraction target through rough measurement of the input image. The window is directly generated based on the original image size using clustering algorithm to avoid the occupation of the main features by invalid information windows. The scanning process of the DF algorithm simplifies the traversal and sampling process of image frames, using effective feature data links to generate intersecting anchor boxes, and quickly locking the effective information window for feature transmission through algorithm training. Finally, the clustering pruning module and joint crossover method are used to achieve rapid output of effective information.

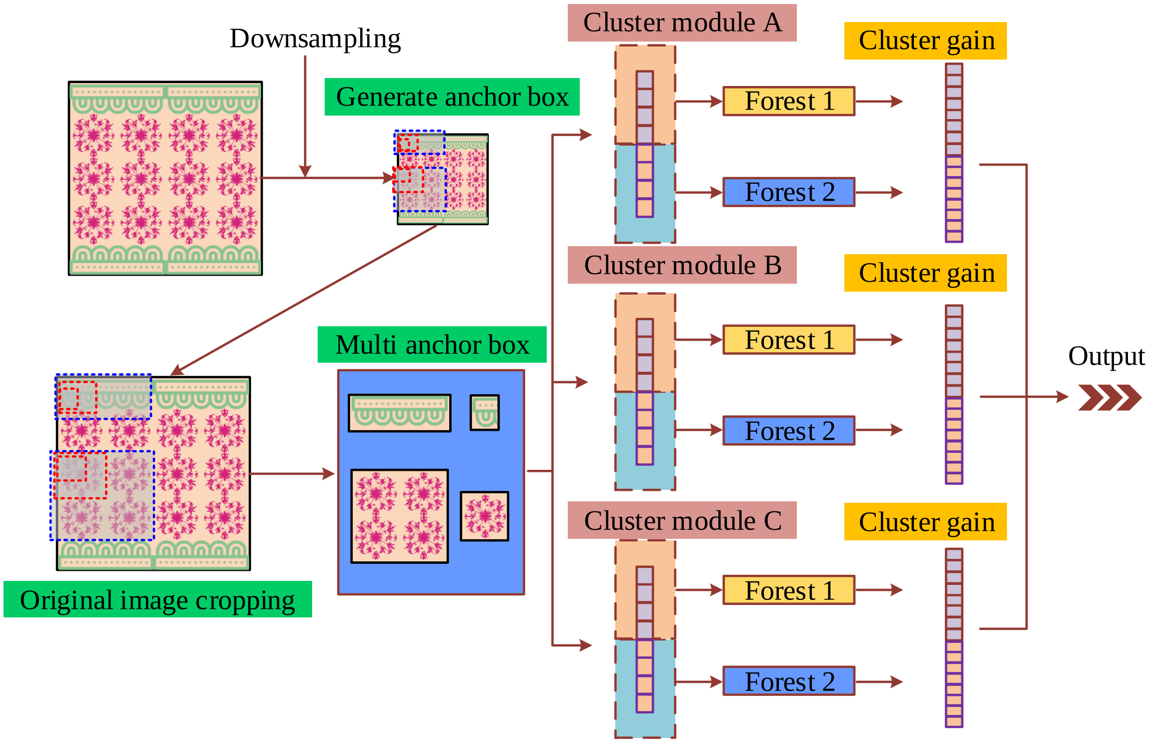


Fig. 3. Prior box workflow.

B. Construction of Ethnic Art Product Image Retrieval Module Combined with CNN

The image retrieval involves users as the main users, so the process inevitably involves personal subjective expression issues [26-27]. There is a semantic gap between the image information subjectively expressed by humans and the actual images recognized by computers. Therefore, the retrieval process needs to convert user expression information into image information and use the computer vision field to process the retrieval problem between images [28-29]. Image retrieval technology based on deep learning has precise and fast processing capabilities for complex and diverse pattern information. The core method of image retrieval technology is the adaptive ability of CNN [30]. Therefore, CNN is combined with DFto jointly construct an image retrieval module for ethnic art design products. The DF-CNN algorithm transmits feature images through sparse connections. The specification of the output feature map is represented by Eq. (7).

$$h_o = \frac{h_{im} + 2 \times padding\_size - k\_h}{stride\_h} \tag{7}$$

In Eq. (7), h_o represents the size of the feature window. h_{im} represents the input image size. $padding\_size$ represents the pixel value of edge extension. k_h signifies the length of the convolution kernel. $stride\_h$ represents the step size. The CNN algorithm achieves image recognition through weight feedback and convolution calculation. The CNN is shown in Fig. 4.

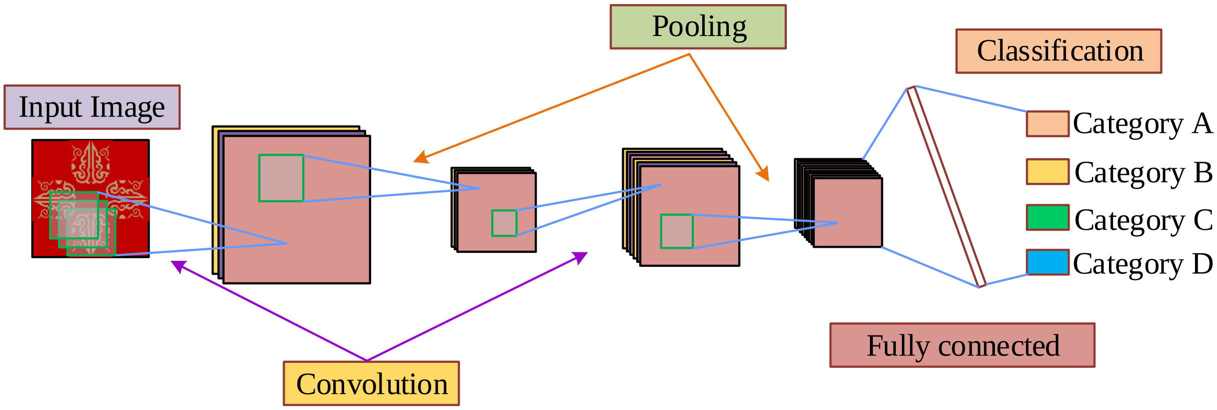


Fig. 4. Convolutional neural network process.

Fig. 4 shows that after the input of the CNN algorithm, the convolutional layers and sub-sampling layers are arranged in an overlapping manner. The collected image features are output through multi-layer convolution and pooling operations. The sub-sampling layer can process and transmit data within the selected range, and the processed features are reduced in model complexity through fuzzy operations. To further enhance the learning ability of the network, ResNet residual network technology is used to complete multi-layer data learning. The calculation process of the residual network is represented by Eq. (8).

$$H\left(x\right) = F(x) + x \tag{8}$$

In Eq. (8), $H\left(x\right)$ represents the functional expression of the residual process. $F(x)$ represents the module expression function in the residual process. x serves as the feature image for the input stage. The residual formula obtained through transformation processing is Eq. (9).

$$y = R(H(x)) \tag{9}$$

In Eq. (9), y represents the output value. R represents the ReLU activation function used in the network model. The residual module can effectively alleviate the gradient vanishing during signal propagation. The study strengthens the application of residual modules by setting a Bottle-neck structure, further increasing the depth of the network and ensuring that the enhanced deep network maintains efficient learning ability. In order to optimize the process of transforming image features from high dimensions to low dimensions, a 1×1 convolutional layer is added to the original CNN model, and local normalization is adopted after each convolutional layer to enhance the data transmission process. The extracted image features are detected by calculating the similarity between the network input and the matched ethnic art patterns using Euclidean distance. The retrieval results are determined based on the similarity measure. The calculation process of Euclidean distance is represented by Eq. (10).

$$d(x,y)\sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{10}$$

In Eq. (10), d represents the feature distance. x_i represents the i -th feature of one of the two images used to calculate distance. x_i represents the i -th feature of another image for calculating distance. Ethnic art products prioritize symmetrical or rotated patterns with simple patterns. The processed patterns are arranged and stacked to form complex patterns with distinctive features. Therefore, in image feature extraction, it is necessary to more effectively identify the basic image patterns. To optimize the visual feature extraction efficiency in image retrieval, the bar attention mechanism is used to process different features. The specific working mode is shown in Fig. 5.

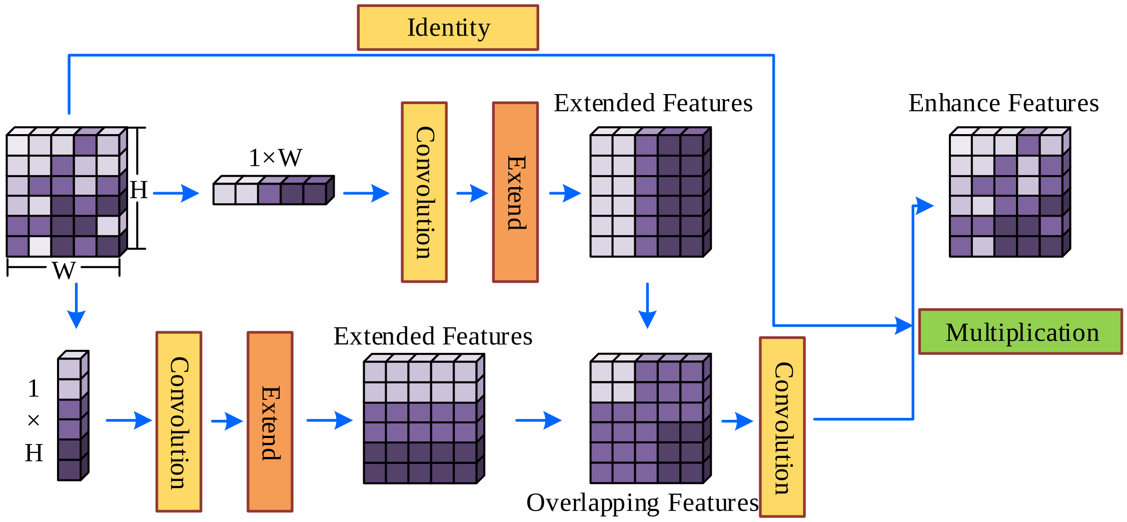


Fig. 5. Bar attention mechanism.

In Fig. 5, the feature extraction mode of the model is composed of a single vertical bar box and a single horizontal bar box. The two sets of standard bar boxes respectively enhance the feature attention points in the current mode, and form a set of enhanced feature data through expansion and combination, filtering out feature maps with less feature information. By optimizing the attention mechanism, the network model magnifies the local feature points of the pattern, resulting in more accurate recognition of effective patterns. The perceptual field of the horizontal bar box in the feature enhancement process is represented by Eq. (11).

$$y_i^h = \frac{1}{W}\sum_{j=1}^{W} x_{i,j} \tag{11}$$

In Eq. (11), y_i^h is the perceived visual field output of the horizontal box. W represents the width value of the space. i signifies the current row value. j signifies the current column value. The perceived visual field of the corresponding vertical bar frame is represented by Eq. (12).

$$y_j^w = \frac{1}{H}\sum_{i=1}^{H} x_{i,j} \tag{12}$$

In Eq. (12), y_j^w D represents the perceived visual field output of the vertical box. H represents the height value of space. Due to the shape setting of the window visual field, a long line connection is established between the discrete region and the main features. The value box of the bar range also makes the feature extraction of the image more targeted, which can enhance attention from two directions and integrate feature maps. To optimize the efficiency of image feature processing, a multi-channel convolution operation is performed, as shown in Fig. 6.

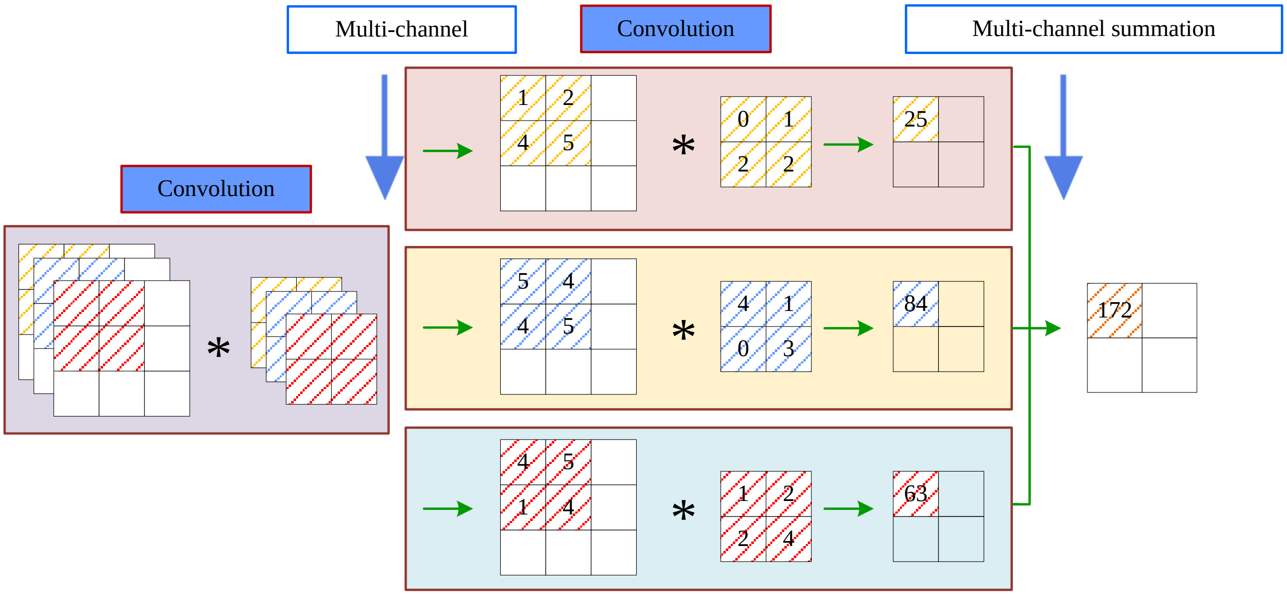


Fig. 6. Multi-channel convolution operation.

In Fig. 6, in the input stage, the convolution kernel dimensionality determines the convolution calculation method. Multiple channels are given priority in completing their respective convolution operations. In the output stage, the sum of convolution values for multiple channels is calculated. The convolution kernel moves with a set stride, and the output size of the feature values is consistent with the dimension size of the convolution kernel. The calculation process of convolution operation is shown in Eq. (13).

$$x^{l+1} = w^l * x^l + b^l \tag{13}$$

In Eq. (13), x^{l+1} represents the output value after convolution operation. x^l represents the operation output of the upper. w^l represents the computed convolution kernel. $*$ represents convolution operation. b^l represents the bias term of the current layer. l represents the number of layers in the convolution operation. To make the self-learning mechanism of the network more efficient, dependency relationships can be established between network channels. The channel attention mechanism can be strengthened through training feedback, which selectively integrates the extracted local features to obtain feature maps with more information. The research model divides adjacent pooling layers using the overlap pooling algorithm, which focuses on global features through the overlapping parts of the region. The specific algorithm process is shown in Fig. 7.

In Fig. 7, it is shown that the feature map after overlapping pooling can cover global features. The input image is convolved and overlapped with pooled feature regions. The overlapped pooled feature map is output using the global average pooling method. The feature maps that have undergone overlapping pooling highlight important features and weaken a small amount of information regions, reducing the image blur problem of global average pooling output. The specification calculation of the new feature map generated by overlap pooling is shown in Eq. (14).

$$n = \frac{m-f}{s} + 1 \tag{14}$$

In Eq. (14), n represents the output value specification. m represents the input value specification. f signifies the size of the convolution kernel. s represents the selected step size. The single channel calculation process of global average pooling is displayed in Eq. (15).

$$\zeta'_n(I) = \frac{\sum_{y_i=1}^{H'}\sum_{x_i=1}^{W'} f(x_i, y_i)}{H' \times W'} \tag{15}$$

In Eq. (15), $\zeta'_n(I)$ is used as the output value of feature map I on channel n. H' represents the height of the newly output feature map. W' represents the width of the newly output feature map. $f(x_i, y_i)$ represents the feature points of the new output feature map. Therefore, when constructing the image retrieval module for ethnic art design products, the DF algorithm is prioritized to enhance feature classification extraction. The DF-CNN is designed in combination with a CNN, and the residual network is used to optimize feature map transmission. The strip attention mechanism and multi-channel operation are used to optimize the image feature extraction process. The global features are integrated through overlap pooling to achieve image retrieval output.

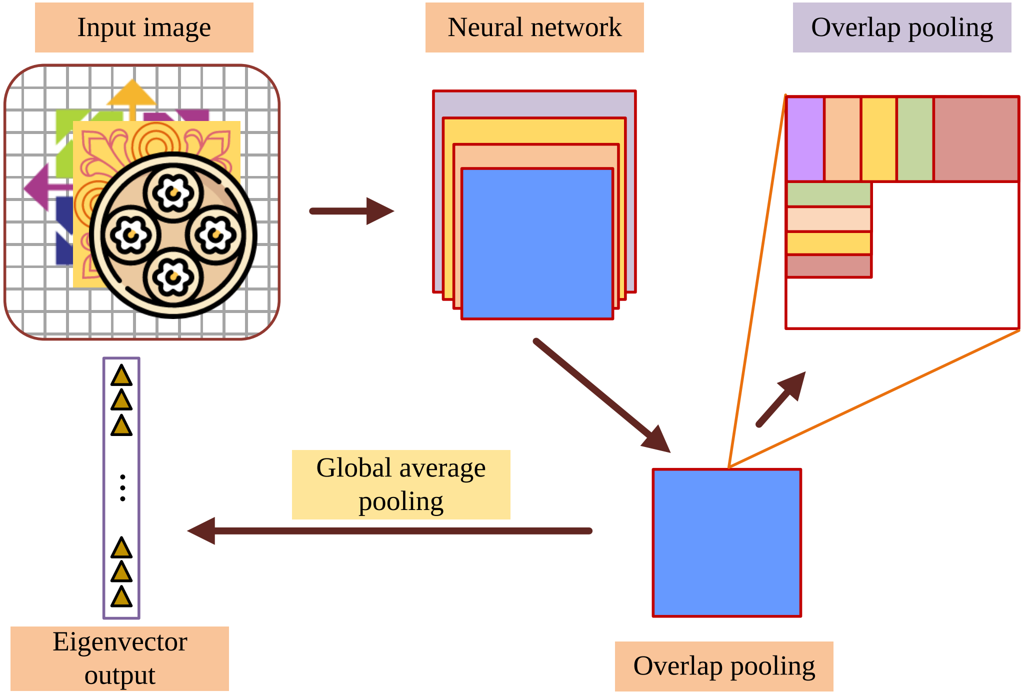
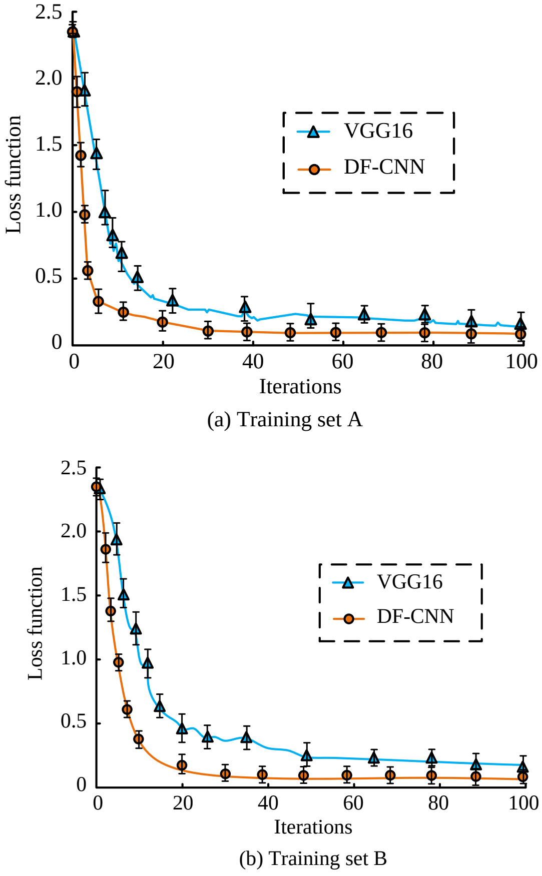


Fig. 7. Overlap pooling and global average pooling.

IV. Results

A. Performance Testing of Image Retrieval Model for Ethnic Art Products

To test the image detection performance of the model, a training set is selected to enhance the learning ability of the model. The pre-testing ratio is set to 10%, and the network threshold of the model is 0.5. The sample images are selected from the CIFAR-10 dataset, and 200 images are set for each group based on the shape of the image pattern as test samples. The convolutional layer during the testing process is set to 16 layers, and the fully connected layer is set to three layers. The learning rate is set to 0.001 and the training iterations are 100. The iterative changes of the loss function in the training set is shown in Fig. 8.

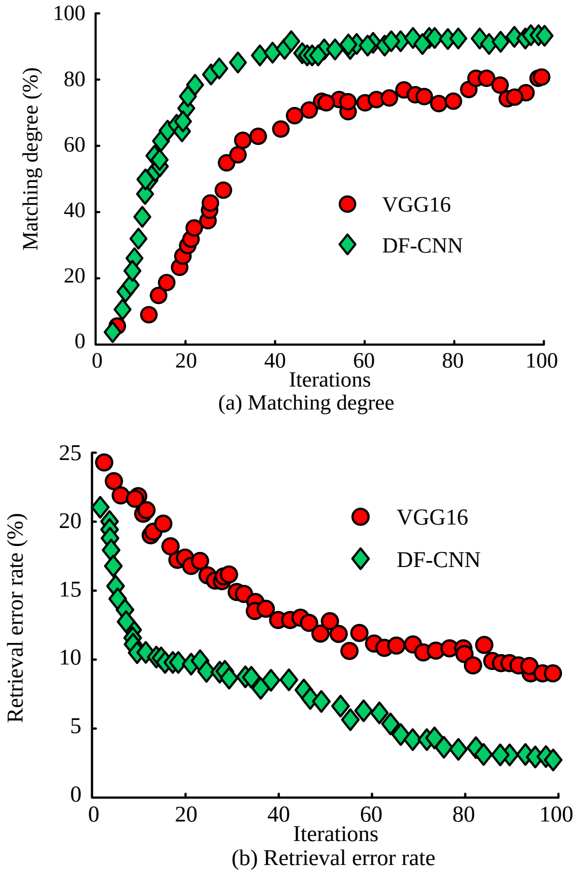


(a) Training set A

(b) Training set B

Fig. 8. Changes in the loss function in the training set.

In Fig. 8 (a), the loss function of the DF-CNN model in training set A decreased with increasing iterations. When the iterations reached 10, it realized stable convergence. The loss function variation of the Visual Geometry Group Network (VGG16) model [31] in training set A also decreased with increasing iterations. When the number of iterations reached 20, the loss function tended to converge, but showed slight fluctuations in subsequent iterations. In Fig. 8 (b), the loss function of the DF-CNN model in training set B was consistent with the results in training set A, but the convergence speed was slightly slower. After 20 iterations, the loss function converged stably. The loss function of VGG16 model in training set B tended to converge after 25 iterations, but the converged loss function value showed a slight rebound. The research model in the training set shows a stable trend and performs better in image detection compared with the VGG16 model. To further validate the pattern retrieval matching of the model in the testing set, the retrieval results of the model for the samples are shown in Fig. 9.



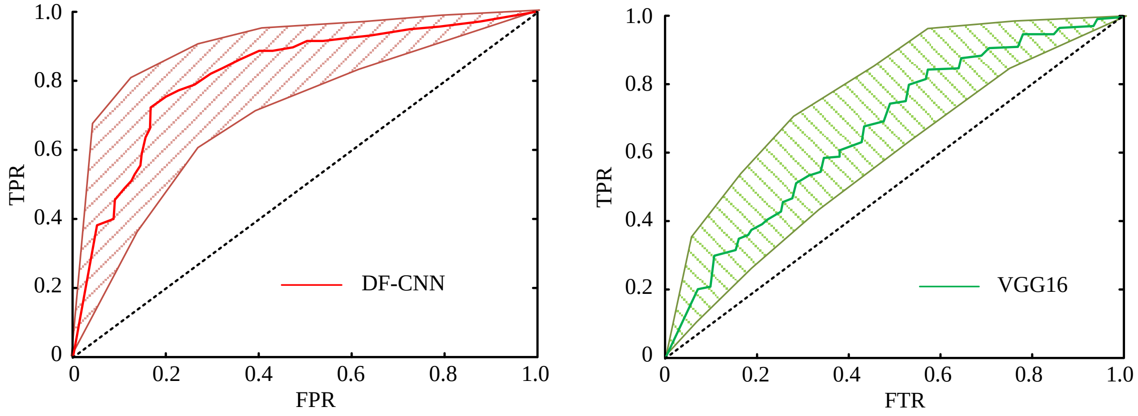
(a) Matching degree

(b) Retrieval error rate

Fig. 9. Pattern retrieval performance in the testing set.

In Fig. 9 (a), the DF-CNN model gradually increased its matching rate with image samples during the iterative learning process. After 23 iterations, the matching degree of the research model reached over 80%. When the model completed learning, the retrieval matching degree of the image reached 91.28%. The VGG16 model had a lower image retrieval matching degree compared with the research model, with an image matching degree of 80.69% after the model completed training and learning. In Fig. 9 (b), the error rate of image retrieval in the DF-CNN model gradually decreased during the iteration process. When the iteration reached 10 times, the image retrieval rate of the model decreased to within 10%. The error rate of the model after training was reduced to 2.91%. The error rate of the VGG16 model in the testing set steadily decreased with increasing iterations, but the error rate optimization effect in the research model during iterations was better than that of the VGG16 model. After 100 iterations of learning, the error rate of the VGG16 model was 9.04%. The trained research model demonstrates good performance in recognizing various types of image patterns, with lower retrieval error rates compared with the VGG16 method, and better image retrieval performance. To ensure the application effectiveness of the research model in image retrieval, the Receiver Operating Characteristic Curve (ROC) is used to evaluate the model.

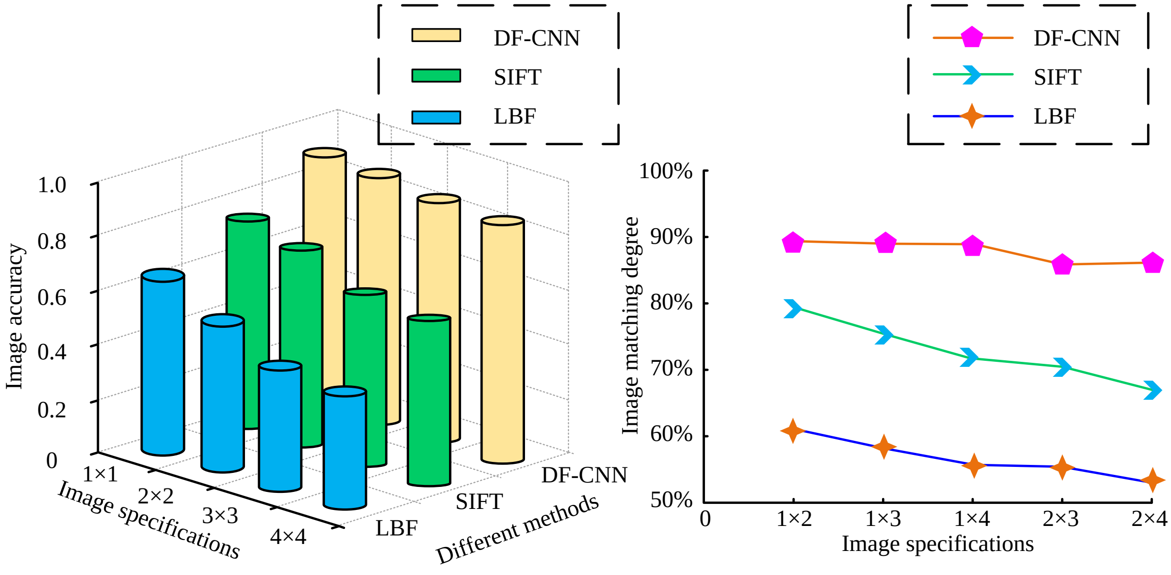
Fig. 10(a) shows the ROC of the DF-CNN model. The ROC curve of the DF-CNN was convex towards the upper left and away from the threshold line, with an Area Under Curve (AUC) of 0.876. Fig. 10(b) shows the ROC curve of the SIFT algorithm. The convex distance of the curve was slightly further to the upper left and the curve range was completely in the lower layer of the research model. The classification performance of the research model performs better throughout the entire process. The ROC curve of the SIFT algorithm was closer to the threshold line and the final AUC value was 0.716. Therefore, the classification performance of the SIFT algorithm [32] is average. This indicates that the image classification performance of the research model has significant advantages, and its ability to retrieve images of ethnic art products is stronger than that of SIFT method in terms of recognition. To further analyze the image comparison ability of the model, the results of non-uniform specification image retrieval under different methods are shown in Fig. 11.



(a) ROC of DF-CNN model

(b) ROC of SIFT model

Fig. 10. Comparison of ROC curves of different models.



(a) Image accuracy

(b) Image matching degree

Fig. 11. Comparison of retrieval effects for multi-specification images.

In Fig. 11 (a), the DF-CNN model exhibits different retrieval accuracy in the selected images of various specifications, with a decreasing trend in retrieval accuracy as the size of the image increases. The retrieval rates of SIFT and LBF algorithms for images of various specifications were consistent with the research method. However, the research model had a relatively small range of accuracy changes for image retrieval of different specifications, with a difference of 4.16% in accuracy under different specifications. The accuracy difference of LBF algorithm and SIFT algorithm for image retrieval of different specifications was 16.21% and 18.97%, respectively. In Fig. 11 (b), the image retrieval accuracy of the three models decreased with increasing image size under irregular image sizes. The DF-CNN model had an average accuracy of 88.49% for non-equal length image type retrieval, while the LBF algorithm and SIFT algorithm had average accuracy of 73.94% and 57.59% for non-equal length image type retrieval, respectively. The results indicate that LBF and SIFT have poor image retrieval performance under the current specifications, while the research model performs well in retrieving images of multiple specifications, which also shows good adaptability in retrieving images with a wide variety of specifications.

B. Application Effect Test of Image Retrieval Model for Ethnic Art Products

To visualize the application effect, a designed image retrieval webpage is used to complete the simulation image retrieval of ethnic art design products. The computer processor used for the simulation experiment is Intel(R) Xeon(R)Platinum, and the GPU specification is RTX 2080 8GB*4. An example of retrieving ethnic art images using research methods is shown in Fig. 12.

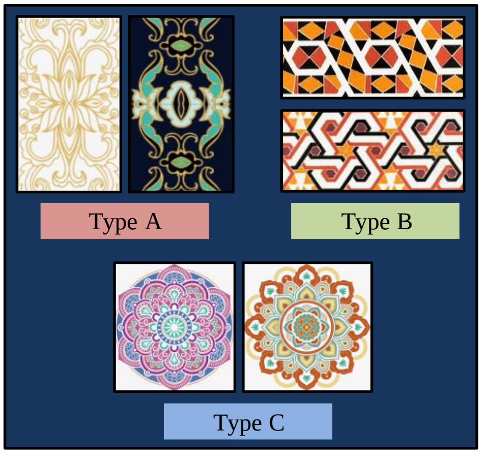
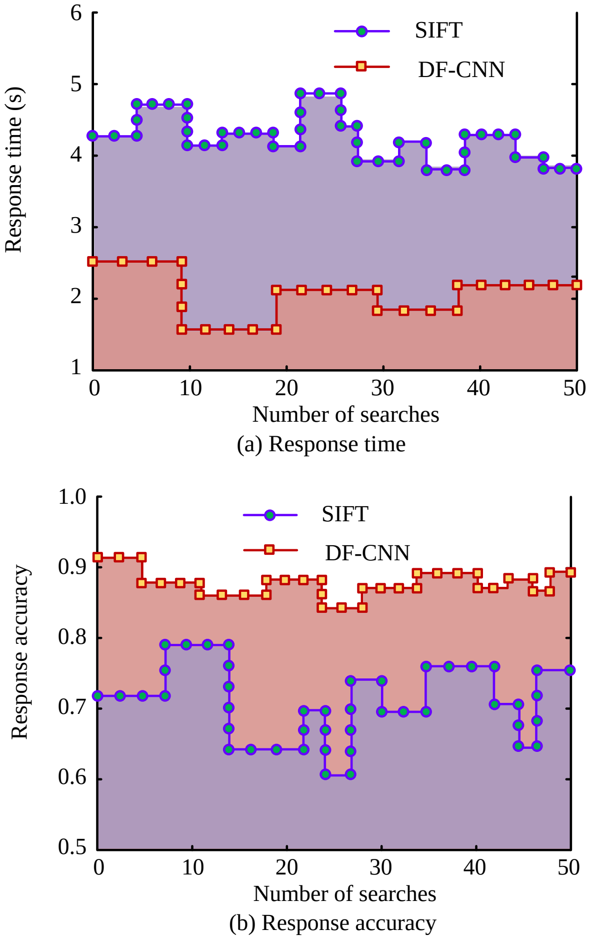


Fig. 12. Retrieve image examples.

Fig. 12 shows the partial image results obtained through keyword retrieval, indicating a clear distinction in the retrieval types of the images. To further test the detection efficiency of the image, the retrieval response time and response accuracy were obtained through 50 retrieval processes, as shown in Fig. 13.

The average response time of the DF-CNN model for 50 image retrieval processes shown in Fig. 13(a) was 1.87s, with the longest response time of the model for retrieval results being 2.47s and the shortest response time being 1.44s. The average response time of the SIFT model for 50 retrieval processes was 4.21s, with the longest response time being 4.86s and the shortest response time being 3.88s. In Fig. 13 (b), the average image retrieval matching degree of the DF-CNN model in 50 searches was 0.879. The optimal matching degree of the research model for images was 0.913, and the matching degree of the image with the worst retrieval effect was 0.843. The average image matching degree of the SIFT model for 50 retrieval processes was 0.701. This indicates that the research model shows a high search speed in the image retrieval, with a response speed improvement of 51.8% compared with the SIFT algorithm. The DF-CNN model also shows a high matching rate for image retrieval, which is 16.8% higher than the image matching rate of the SIFT algorithm. To compare the image retrieval performance under various conditions, the image retrieval results of the research model within one week are summarized in Table I.



(a) Response time

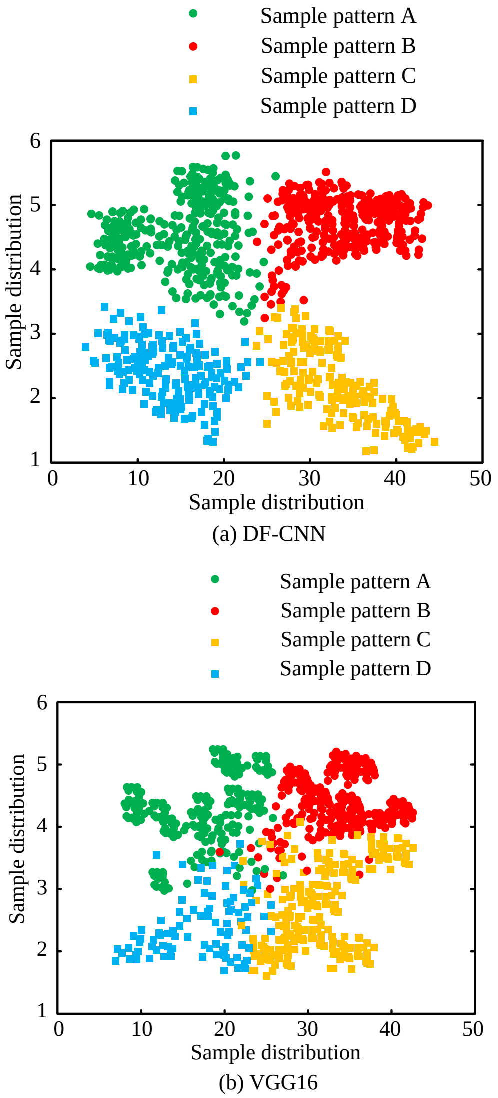
(b) Response accuracy

Fig. 13. Search response during the application process.

In Table I, the three types of models with a code length of 128 had shorter response times for image detection, among which the DF-CNN model had the fastest response speed in the detection process. The three types of models with a code length of 512 showed longer response times for image detection, and the DF-CNN model remained the fastest response speed. However, as the code length increased, the mean Average Precision (mAP) and recall of the three types of models showed an upward trend. The DF-CNN model demonstrated good retrieval precision and sensitivity. Compared with the VGG16 algorithm and DF-CNN algorithm, the mAP value optimization range of the research model was 6.1% -29.8%, and the sensitivity optimization range was 15.33% -36.9%. The research model shows that the retrieval speed of images is faster under shorter code lengths, but the retrieval precision is correspondingly reduced. The retrieval precision of images is improved under longer code lengths, but the corresponding retrieval time is longer. The DF-CNN model shows excellent performance in both detection precision and response time during the detection process. To distinguish and recognize different patterns, the clustering results of detection samples under different algorithms are compared, as shown in Fig. 14.

TABLE I. COMPARISON OF RETRIEVAL PERFORMANCE OF DIFFERENT MODELS

Retrieval Model	Code-Length	Recall	mAP	Time/s	References
DF-CNN	128	53.30%	76.4%	1.64	/
DF-CNN	512	57.40%	90.1%	4.62	/
VGG16	128	46.80%	69.3%	1.82	[31]
VGG16	512	48.60%	84.6%	5.07	[31]
SIFT	128	33.90%	45.3%	2.85	[32]
SIFT	512	36.20%	63.2%	8.74	[32]

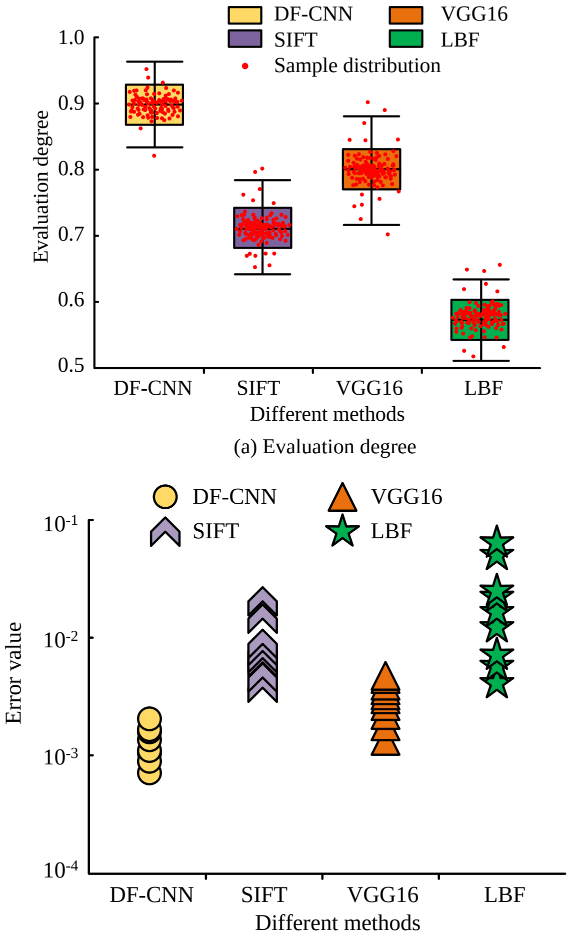


(a) DF-CNN

(b) VGG16

Fig. 14. Clustering of detection samples under different algorithms.

In Fig. 14 (a), the four types of pattern detection samples in the DF-CNN model were clustered separately, and the boundary parts of the four types of samples were adjacent but not intersecting. In Fig. 14 (b), the four pattern type detection samples in the VGG16 model also exhibited their own clustering, but there was partial intersection at the boundaries of the four types of samples. The DF-CNN model showed good recognition performance for four patterns and strong recognition ability for similar images. The VGG16 model can achieve basic recognition and detection for four types of patterns. When the patterns presented in the retrieved images are similar, the VGG16 model may exhibit confusion in image detection. The patterns of ethnic art products may appear similar but not identical, and the VGG16 model shows a significant lack of retrieval ability for such images. The DF-CNN model can accurately identify the differences in such images, indicating that the research model has a better retrieval application effect for ethnic art products. To verify the detection accuracy of the model, the multi-sample image detection accuracy and image detection error values under different models are shown in Fig. 15.



(a) Evaluation degree

(b) Error range

Fig. 15. Image detection accuracy and detection error.

In Fig. 15 (a), the average image accuracy evaluation value of the DF-CNN model in multi-sample image detection was 0.898, with only a few samples scattered outside the detection floating range. The image accuracy evaluation values of SIFT algorithm, VGG16 algorithm, and LBF algorithm were 0.731, 0.804, and 0.579, respectively, and all three algorithms had some samples scattered outside the detection floating range during the image detection process. In Fig. 15 (b), the detection error range of the DF-CNN model in multi-sample image detection was 8.7×10-4-4.3×10-3, while the image detection error ranges of the SIFT algorithm, VGG16 algorithm, and LBF algorithm were 4.7×10-3-4.1×10-2, 1.1×10-3-7.9×10-3, and 4.8×10-3-9.1×10-2, respectively. This demonstrates that the research method has good stability in image detection function. Compared with other algorithms, the image accuracy has always been maintained at a good level. The image detection error of the research model is smaller and the error value is lower compared with other algorithms, indicating that the research model has good image detection ability during operation.

C. Discussion

The above research results show that the research method has good matching performance for the retrieval of images of ethnic art products, and the accuracy of image retrieval is better than that of the SIFT algorithm. Due to the SIFT algorithm's emphasis on local features in image detection, and the lack of significant differences in local features of ethnic art products, the retrieval matching degree of images decreases. The research model utilizes the cascaded structure of the DF algorithm to enhance the depth of image feature mining, strengthen the correlation between retrieval keywords and corresponding images, and thus improve the retrieval accuracy of the design module. At the same time, the retrieval efficiency of the research method also shows significant advantages. In the process of retrieving ethnic art images, the retrieval response time of the research method is significantly better than that of the VGG16 algorithm. The VGG16 algorithm enhances its feature extraction ability through multi-layer convolution, but due to the complex calculation process, the retrieval speed of the algorithm decreases. The prior box design of the research model increases the pre-screening of image features, reduces the computational complexity of the image-matching process, and thus achieves an improvement in retrieval efficiency. It can be seen that the research model achieves synchronous improvement in retrieval matching and retrieval efficiency through the fusion of advantages between different algorithms.

V. Conclusion

In order to enhance the retrieval ability of art galleries for ethnic art patterns, an image retrieval model is constructed by combining the DF algorithm with CNN. The decision tree was used to explore the transformation paths of images, and the overlap window was used to enhance the global features of features. Multi-granularity scanning processes optimized feature loss during dimension transformation. The Euclidean distance algorithm was used to calculate the similarity of image features. Parallel multi-channel convolution was applied to optimize the transmission rate of features, focusing on global features through overlapping pooling algorithm, and completing image retrieval output by combining global and local features. The research results indicated that the error rate of the DF-CNN model was reduced to 2.91% after training, which was 6.13% lower than the retrieval error rate of the VGG16. The difference in retrieval accuracy of the DF-CNN model for multi-specification image retrieval was 4.16%, which was 12.05%-14.81% higher than other algorithms, indicating that the research model had good adaptability to retrieval of various image types and stable retrieval performance for images. In the application process, the average retrieval response time of the DF-CNN model was 1.87s, and the longest response time was 2.47s, which was 51.8% higher than the response rate of the SIFT algorithm. The research model had good image retrieval accuracy while maintaining a high response rate. The average image retrieval matching degree of the model was 0.879, which was 16.8% higher than the SIFT algorithm. The application effect of the research model in practical processes is good, and the efficiency of image retrieval is high. With the widespread dissemination of ethnic art products in more fields, they have also been applied in virtual scenes. However, in the practical application of retrieval methods, virtual scenes are often not included in the retrieval scope, resulting in a lack of retrieval data for virtual scenes. At the same time, the research method has increased the retrieval ability of ethnic art images by adding multiple complex graphic processing layers, which has resulted in a decrease in the retrieval speed of simple images. Therefore, in future research plans, image retrieval scenarios can be expanded to achieve further expansion of retrieval databases. It is possible to attempt to classify the retrieval paths of images, achieving hierarchical processing of simple and complex images, thereby further improving the efficiency of image retrieval and meeting the retrieval needs of different users.

References

[1] C. A. Burks, T. I. Russell, D. Goss, D. Goss, G. Ortega, and G. W. Randolph, "Strategies to increase racial and ethnic diversity in the surgical workforce: A state of the art review," Otolaryng. Head Neck, vol. 166, no. 6, pp. 1182-1191, April, 2022.

[2] A. Olivares and J. Piatak, "Exhibiting inclusion: An examination of race, ethnicity, and museum participation," Int. J. Voluntary Nonprofit Organ., vol. 33, no. 1, pp. 121-133, February, 2022.

[3] T. Zhang, B. Li, and N. Hua, "Chinese cultural theme parks: text mining and sentiment analysis," J. Tour. Cult. Change, vol. 20, no. 1-2, pp. 37-57, January, 2022.

[4] N. Arora and S. C. Sharma, "ETLBP and ERDLBP descriptors for efficient facial image retrieval in CBIR systems," Multimedia Tools Appl., vol. 83, no. 4, pp. 9817-9851, June, 2024.

[5] M. Majhi, A. K. Pal, J. Pradhan, S. H. Islam, and M. K. Khan, "Computational intelligence based secure three-party CBIR scheme for medical data for cloud-assisted healthcare applications," Multimedia Tools Appl., vol. 81, no. 29, pp. 41545-41577, February, 2022.

[6] S. Saurav, R. Saini, and S. Singh, "Fast facial expression recognition using boosted histogram of oriented gradient (BHOG) features," Pattern Anal. Appl., vol. 26, no. 1, pp. 381-402, September, 2023.

[7] S. Suganyadevi, V. Seethalakshmi, and K. Balasamy, "A review on deep learning in medical image analysis," Int. J. Multimed. Inf. R., vol. 11, no. 1, pp. 19-38, September, 2022.

[8] M. K. Kelishadrokhi, M.Ghattaei, and S. Fekri-Ershad, "Innovative local texture descriptor in joint of human-based color features for content-based image retrieval," Signal, Image Video P., vol. 17, no. 8, pp. 4009-4017, July, 2023.

[9] N. Keisham and A. Neelima, "Efficient content-based image retrieval using deep search and rescue algorithm," Soft Comput., vol. 26, no. 4, pp. 1597-1616, January, 2022.

[10] W. Wang, P. Jiao, H. Liu, X. Ma, and Z. Shang, "Two-stage content based image retrieval using sparse representation and feature fusion," Multimed. Tools Appl., vol. 81, no. 12, pp. 16621-16644, March, 2022.

[11] C. Ning, Y. Di, and L. Menglu, "Survey on clothing image retrieval with cross-domain," Complex Intell. Syst., vol. 8, no. 6, pp. 5531-5544, May, 2022.

[12] H. Zhuang, X. Liu, Y. Yan, D. Zhang, J. He, J. He, X. Zhang, H. Zhang, and M. Li, "Integrating a deep forest algorithm with vector-based cellular automata for urban land change simulation," Trans. GIS, vol. 26, no. 4, pp. 2056-2080, April, 2022.

[13] A. Hamedianfar, C. Mohamedou, A. Kangas, and J. Vauhkonen, "Deep learning for forest inventory and planning: A critical review on the remote

sensing approaches so far and prospects for further applications," Forestry, vol. 95, no. 4, pp. 451-465, October, 2022.

[14] M. A.Shaaban, Y. F. Hassan, and S. K. Guirguis, "Deep convolutional forest: a dynamic deep ensemble approach for spam detection in text," Complex Intell. Syst., vol. 8, no. 6, pp. 4897-4909, April, 2022.

[15] T. Huang, Q. Zhang, X. Tang, S. Zhao, and X. Lu, "A novel fault diagnosis method based on CNN and LSTM and its application in fault diagnosis for complex systems," Artifi. Intell. Rev. vol. 55, no. 2, pp. 1289-1315, April, 2022.

[16] A. Tayal, J. Gupta, A. Solanki, K. Bisht, A. Nayyar, and M. Masud, "DL-CNN-based approach with image processing techniques for diagnosis of retinal diseases," Multimedia Syst., vol. 28, no. 4, pp. 1417-1438, March, 2022.

[17] M. Zhitomirsky-Geffet and S. Minster, "Cultural information bubbles: A new approach for automatic ethical evaluation of digital artwork collections based on Wikidata," Digit. Scholarsh. Hum., vol. 38, no. 2, pp. 891-911, June, 2023.

[18] W. Serhan, "Symbolic capital and the inclusion of ethnic minority artists in Dublin and Warsaw," Ethnicities, vol. 24, no. 3, pp. 475-496, June, 2024.

[19] M. Zhitomirsky-Geffet, I. izhner, and S. Minster, "What do they make us see: a comparative study of cultural bias in online databases of two large museums," J. Doc., vol. 79, no. 2, pp. 320-340, July, 2023.

[20] C. Sutherland, "Transcending ethnicity through photography: representing the Cham," Asian Ethn., vol. 23, no. 1, pp. 127-145, March, 2022.

[21] J. Guo, Y. Cai, Y. Fan, F. Sun, R. Zhang, and X. Cheng, "Semantic models for the first-stage retrieval: A comprehensive review," ACM Trans. Inform. Syst., vol. 40, no. 4, pp. 1-42, March, 2022.

[22] Y. Cheng, X. Zhu, J. Qian, F. Wen, and P. Liu, "Cross-modal graph matching network for image-text retrieval," ACM Trans. Multim. Comput., vol. 18, no. 4, pp. 1-23, March, 2022.

[23] N. T. Dinh, N. T. U. Nhi, T. M. Le, and T. T. Van, "A model of image retrieval based on KD-tree random forest," Data Technol. Appl., vol. 57, no. 4, pp. 514-536, May, 2023.

[24] N. Bharatha Devi, "Satellite image retrieval of random forest (rf-PNN) based probablistic neural network," Earth Sci. Inform., vol. 15, no. 2, pp. 941-949, February, 2022.

[25] P. Ma, Y. Wu, Y. Li, L. Guo, H. Jiang, and X. Zhu, "HW-Forest: Deep forest with hashing screening and window screening," ACM Trans. Knowle. Discov. Data, vol. 16, no. 6, pp. 1-24, July, 2022.

[26] S. Kumar, A. K. Pal, S. K. H. Islam, and M. Hammoudeh, "Secure and efficient image retrieval through invariant features selection in insecure cloud environments," Neural Comput. Appl., vol. 35, no. 7, pp. 4855-4880, June, 2023.

[27] S. Heller, V. Gsteiger, W. Bailer, C. Gurrin, B. Þ.Jónsson, and J. Lokoč, "Interactive video retrieval evaluation at a distance: comparing sixteen interactive video search systems in a remote setting at the 10th Video Browser Showdown," Int. J. Multimed. Inf. Retr., vol. 11, no. 1, pp. 1-18, January, 2022.

[28] L. Shi, J. Du, G. Cheng, X. Liu, Z. Xiong, and J. Luo, "Cross-media search method based on complementary attention and generative adversarial network for social networks," Int. J. Intell. Syst., vol. 37, no. 8, pp. 4393-4416, November, 2022.

[29] J. Guo, Y. Cai, Y. Fan, F. Sun, R. Zhang, and X. Cheng, "Semantic models for the first-stage retrieval: A comprehensive review," ACM Trans. Inform. Syst., vol. 40, no. 4, pp. 1-42, March, 2022.

[30] P. Preethi and H. R. Mamatha, "Region-based convolutional neural network for segmenting text in epigraphical images," Artif. Intell. Appl., vol. 1, no. 2, pp. 119-127, September, 2023.

[31] S. Kumar, M. K. Singh, and M. Mishra, "Efficient deep feature based semantic image retrieval," Neural Process. Lett., vol. 55, no. 3, pp. 2225-2248, January, 2023.

[32] W. Wang, P. Jiao, H. Liu, X. Ma, and Z. Shang, "Two-stage content based image retrieval using sparse representation and feature fusion," Multimed. Tools Appl., vol. 81, no. 12, pp. 16621-16644, March, 2022.